ENGINEERING ELONGATION FACTOR TU AND TRNAS TO BETTER ACCOMMODATE NON-CANONICAL AMINO ACIDS DURING TRANSLATION

A Dissertation
Presented to
The Academic Faculty

by

Vanessa E. DeLey Cox

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2018

ENGINEERING ELONGATION FACTOR TU AND TRNAS TO BETTER ACCOMMODATE NON-CANONICAL AMINO ACIDS DURING TRANSLATION

Approved by:

Dr. Eric Gaucher, Co-Advisor
School of Biology
Georgia Institute of Technology

Dr. M.G. Finn
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren Williams, Co-Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Nicholas Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Adegboyega Oyelere
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: July 23, 2018

To RC, TD, JC, and all my friends on the 5th

ACKNOWLEDGEMENTS

I would like to acknowledge my dissertation committee: Drs. M.G. Finn, Loren Williams, Nicholas Hud, Adegboyega Oyelere, and Eric Gaucher. In particular, Dr. Oyelere offered advice on writing a manuscript, and Dr. Williams agreed to serve as co-advisor. Drs. Finn, Mary Peek, Hui Zhu, and Kenyetta Johnson graciously assisted with securing funding. My supervisor, Dr. Gaucher generously offered funding and resources which ultimately lead to a publication; he also allowed me to author a review article and a book chapter for which I am truly appreciative. Additional technical support and supervision came from the Petit Institute for Bioengineering and Bioscience's Systems Mass Spectrometry Core and the Office of Radiological Safety. Additional funding came from the NIH, ACS, and ASCB.

I would like to recognize the Gaucher Lab with special mention of Dr. Megan Cole who generated the consensus sequence and REAP library. I would like to extend my deep gratitude to the fifth floor of EBB for their support with both the manuscript and lab work, including Drs. Corey Wilson, Julie Champion, Andreas Bommarius, Ravi Kane, and Donald Doyle; the entire Wilson and Kane labs; Drs. James Davey, John Robbins, and Harrison Rose; and Mr. Ammar Arsiwala and Ms. Jennifer Farrar. For their counsel and support, I credit Drs. Seth Marder and Erik Wasinger. This section could not be complete without mention of Margaret Korte, her high standards, and unyielding determination. Finally, I would like to thank my family for their consistent support and understanding. If I forgot anyone, please accept my sincerest apologies; the fault is mine and I would gladly buy you a cup of coffee.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

A, Ala	alanine
aa	amino acid
aaRS	aminoacyl-tRNA synthetase
aa-tRNA	aminoacyl-tRNA
ASR	ancestral sequence reconstruction
ATP	adenosine triphosphate
C, Cys	cysteine
CAT	chloramphenicol acetyltransferase
Ci	curie (as unit)
Cmp	chloramphenicol
CPM	counts per minute
D, Asp	aspartic acid
dATP	deoxyadenosine triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E, Glu	glutamic acid
EF-	elongation factor variant
EF-Tu	elongation factor Tu
F, Phe	phenylalanine
fMet	formyl-methionine
g	gram
G, Gly	glycine

H, His	histidine
H_2O	water
H_6	hexahistidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
I, Ile	isoleucine
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl-β-D-thiogalactopyranoside
K, Lys	lysine
Kan	kanamycin
L	liter (as unit)
L, Leu	leucine
m	milli (as unit)
M	molar (as unit)
M, Met	methionine
$MgCl_2$	magnesium chloride
N, Asn	asparagine
NaCl	sodium chloride
NaOAc	sodium acetate (solution)
ncAA	non-canonical amino acid
OD_{600}	optical density measured at 600nm
OTS	orthogonal translation system
P, Pro	proline
PPase	pyrophosphatase
Q, Gln	glutamine
R, Arg	arginine

REAP	reconstructing evolutionary adaptive paths
RNA	ribonucleic acid
S, Ser	Serine
Sec	selenocysteine
Sep	phosphoserine
T, Thr	threonine
Tet	tetracycline
tRNA	transfer-RNA
V, Val	valine
W, Trp	tryptophan
WT	wild type
Y, Tyr	tyrosine
μ	micro (as unit)

SUMMARY

Genetic code expansion is a central goal of protein research and engineering with a broad range of applications. The ability to reliably incorporate non-canonical amino acids (ncAAs) in a site-specific manner has expanded the protein engineering toolbox to enable the functionalization of proteins with affinity, spectroscopic, and chemically functional components.[1] Consequently, incorporation of functional ncAAs has enabled bio-orthogonal modification of proteins, a powerful and emerging tool critical to the development of both fundamental protein science and applied biotechnologies.[2] The most common technique for the translation of proteins incorporating site-specific ncAA mutations is amber codon suppression.[3] This technique leverages an orthogonal translation system (OTS) consisting of a dedicated aminoacyl-tRNA synthetase (aaRS):tRNA pair mediating the translation of a protein with a specific ncAA at the target position encoded by a repurposed amber codon.[4] Typically, ncAA co-translation in *Escherichia coli* is mediated by an OTS developed from an engineered bio-orthogonal tyrosyl-tRNA synthetase:tRNA pair from *Methanococcus jannaschii*.[1] However, despite the potential promised by an expanded genetic code, the routine application of the OTS strategy is consistently hindered by considerable and recurring barriers.[5]

These persistent challenges include cross-reactive OTSs, incompatibility with endogenous elongation factors, and discrimination by additional translation components. These factors affect both the yield and purity of co-translated ncAA proteins as well as the fitness and viability of the host microorganism.[6-10] Furthermore, the diversity of these modifications is reduced to a specific set of ncAAs compatible with existing and

engineered translation machinery, thereby significantly reducing the readily available scope of potential chemistries and applications. Consequently, current efforts are commonly limited to single-site incorporation with a reduced set of ncAAs.[11] These challenges highlight a pressing need to pursue improved engineering strategies beyond OTS development that will enable translation of increasingly complex peptide products, with multi-site incorporation and multiple ncAA incorporation.[12, 13]

One obstacle limiting expansion of the genetic code is elongation factor Tu (EF-Tu), a guanosine triphosphatase (GTPase). EF-Tu serves two functions in translation. While most commonly recognized for translocation of aa-tRNAs to the ribosome, it also plays a critical role in quality control.[14] All twenty aa-tRNAs associate with EF-Tu having carefully tuned interactions that prevent misacylated tRNAs from being delivered to the ribosome for translation. Similar to misacylated tRNAs, ncaa-tRNAs are non-native substrates and can be discriminated by EF-Tu thus preventing their incorporation into the translated protein. Past efforts have typically circumvented EF-Tu's editing mechanism by targeting ncAAs that are tolerated as substrates. For particularly intractable ncAAs, orthogonal EF-Tus have been developed or tRNAs have been engineered to optimize their binding with EF-Tu.[15] However, these efforts fail to recognize EF-Tu's comprehensive effect on translation.[11] As a result, engineering EF-Tu to accept an expanded set of ncaa-tRNA substrates represents a unique opportunity for expanding ncAA incorporation.[11, 15-18]

We hypothesize that EF-Tu's rejection of ncaa-tRNAs is a significant limiting factor in the translation of ncAAs. Using both *in vitro* and *in vivo* assays, we have evaluated tRNA and EF-Tu variants' ability to efficiently translate ncAAs, specifically valine analogs

(*in vitro*) and phosphoserine (*in vivo*). Broadly, the research presented herein works toward characterizing the tRNA and EF-Tu mutations required to overcome a limiting step in the translation of ncAAs and promote more efficient delivery of ncAA-tRNAs to the ribosome. In contrast to typical efforts, we demonstrate the advantages of using computational analysis to broaden EF-Tu's substrate specificity. By improving host organism fitness and potentially allowing researchers to circumvent the question of ncAA compatibility with EF-Tu, this approach directly complements current research demonstrating the importance of EF-Tu engineering for the achievement of future goals further expanding the genetic code.

CHAPTER 1. INTRODUCTION

1.1 Abstract

In recent years, ncAAs have established their importance in a wide range of fields from pharmaceuticals to polymer science. NcAAs can add any number of chemical groups to proteins for expanded and enhanced functionality, yet despite the importance of ncAAs in research, scientists' ability to utilize these unique building blocks has been limited by their ability to incorporate them efficiently into proteins. To address this challenge, researchers have examined how the canonical twenty amino acids are incorporated, regulated, and modified in nature. This chapter focuses on achievements and techniques used to engineer the ribosomal translation machinery including introducing orthogonal translation components, directed evolution, and computationally-derived ancestral protein sequences.

1.2 Introduction

The cellular translation machinery has evolved to translate a specific set of twenty amino acids. Despite the limited variety of chemical functional groups offered by the canonical amino acids, all organisms utilize the same set of twenty to build proteins for a tremendous variety of cellular applications. One might question how the wide variety of chemical functionality required of proteins *in vivo* is even possible when working with such a restricted set of building blocks. Nature has addressed this challenge by expanding the chemical complexity of protein chemistry through extensive use of post-translational modifications and cofactors, but these cellular techniques can be difficult to harness in the

laboratory. One method for researchers to expand or manipulate protein chemistry is to directly incorporate chemical functionality during translation using ncAAs, meaning amino acids other than the canonical twenty. However, ncAAs are frequently rejected by the translation machinery and suffer from low translation rates. Other challenges include incorporation of an ncAA at multiple positions in the same protein or translating several different ncAAs into a single protein. However, despite the challenges associated with using ncAAs, many scientific fields are actively benefiting from their application.

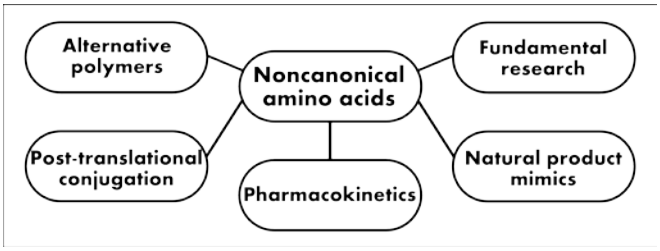

Figure 1.1 Applications of ncAAs. A web diagram illustrating research areas that employ ncAAs.

NcAAs offer a wide variety of chemical functionalities that researchers can exploit for any number of uses such as chemical tags, enhanced or altered functionality, increased stability, or post-translational modifications (Figure 1.1). For example, biotherapeutics depend on ncAAs to address a variety of challenges including low permeability across biological barriers and rapid degradation *in vivo*.[19, 20] NcAAs are also used to chemically tag proteins at a specific location without labeling non-target amino acids or proteins. Such tagged proteins have application in a range of fields including *in vivo* cell imaging, proteomics, and drug-delivery systems.[21-24] Translation of ncAAs is also used in basic protein research to directly incorporate post-translational modifications during translation, thereby offering reliable protein modification while avoiding the challenges associated with post-translational enzymatic or chemical methods.[17, 25] Even nanoscience and polymer

science are exploring the expanded possibilities ncAAs offer.[26] Due to their many applications, researchers have sought methods to insert ncAAs into peptides and proteins.

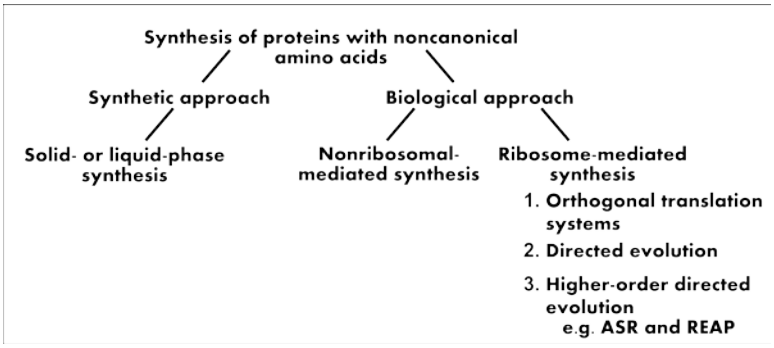

Figure 1.2 Techniques summary. This flow chart shows methods used to incorporate ncAAs into peptides and proteins. This chapter focuses on methods highlighted under "ribosome-mediated synthesis" wherein the ribosome translation machinery is engineered to promote efficient translation of ncAAs.

A number of techniques exist to facilitate ncAA incorporation into proteins and peptides, including synthetic methods such as solid-phase synthesis, *in vitro* systems, and biological methods.[27-29] Biological approaches to incorporating ncAAs are split into non-ribosomal protein synthesis and ribosome-mediate translation. Herein we will focus on the latter (Figure 1.2).[30] In nature, ribosome-mediated translation requires many cellular components including the ribosome, transfer-RNAs (tRNAs), aminoacyl-tRNA synthetases, and elongation factors which synthesize proteins *in vivo* (Figure 1.3). Engineering the ribosome translation machinery to translate ncAAs has required overcoming number of biological proofreading steps that reject ncAAs during translation. Researchers have addressed these challenges using a variety of techniques including developing orthogonal translation components, directed evolution of translation factors, and ancestral sequence reconstruction.

1.3 Engineering translation machinery

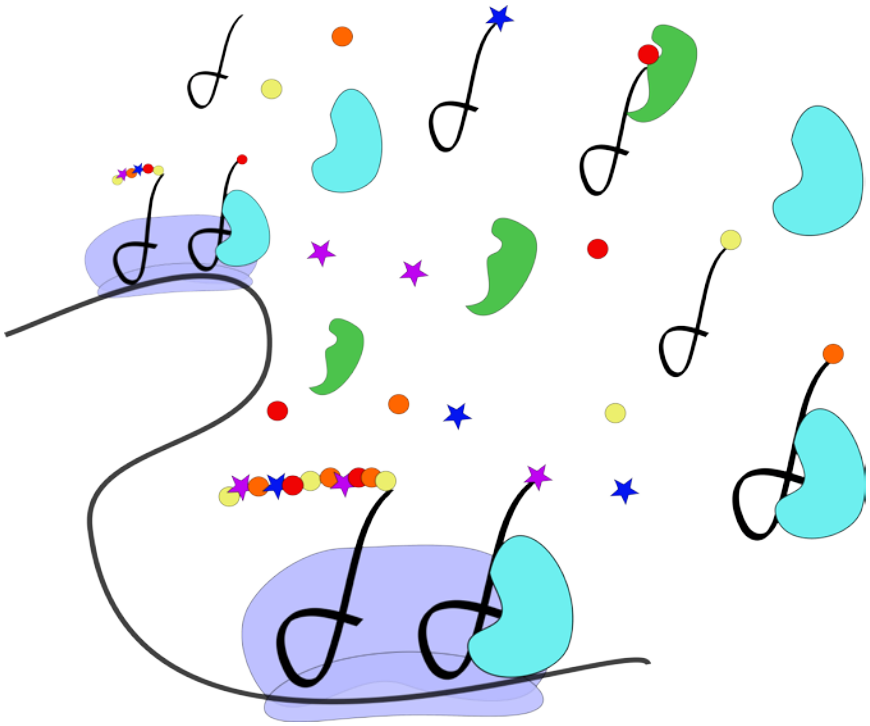

Figure 1.3 Cartoon of translation machinery. Ribosome (lavender), EF-Tu (aqua), synthetases (green), tRNA (black), amino acids (circles) and ncAA (stars) are represented.

Cellular translation utilizes a number of proofreading steps during which it can identify and reject ncAAs, refusing to translate them into a protein. Specific steps in translation allow the ribosome, tRNAs, synthetases, and elongation factor Tu (EF-Tu) to ensure accurate translation of the canonical amino acids and exclusion of any unnatural ones. In the first proofreading step, the synthetase acylates the corresponding tRNA with its correct amino acid. Next, the EF-Tu double checks the aminoacyl-tRNA is charged with the correct amino acid and delivers the complex to the ribosome for translation. Finally, the ribosome may fail accommodate the ncAA and may refuse to incorporate it into a protein. Each of these three steps is a point at which an ncAA may be rejected by the translation machinery. Researchers are working to address each of these proofreading steps in translation.

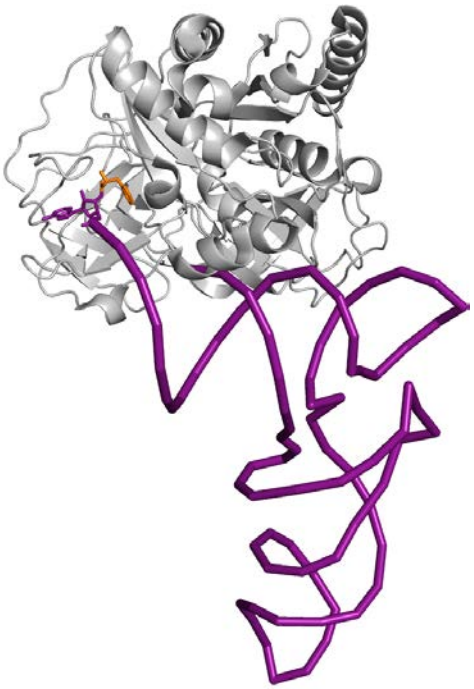

Figure 1.4 EF-Tu crystal structure (gray) complexed with tRNAPhe (purple) and the amino acid target (orange).

Due to research performed in this area, the ribosome is known to be relatively promiscuous and can often accommodate ncAAs; however synthetases, tRNAs, and EF-Tus often contribute to low rates of ncAA incorporation.[17, 31, 32] Researchers have developed methods to engineer these components of the translation machinery. Their techniques include using orthogonal translation components, directed evolution of proteins, and ancestral proteins.

Great strides have been made by researchers to address tRNAs' and synthetases' exclusion of ncAAs. Synthetases and tRNAs work in cellular translation as an orthogonal pair, specific for an amino acid be it canonical or unnatural. To incorporate an ncAA, a new, orthogonal tRNA and synthetase are required.[3, 33] This means that although the tRNA and synthetase must recognize each other, it is critical they do not interact with any other

tRNAs or synthetases in the cell since these pairs are already assigned to a specific amino acid. Rather than develop tRNAs or synthetases *de novo*, a common approach is to adopt a tRNA/synthetase pair from another domain of life. Since tRNAs and synthetases are often domain or even species specific, a pair from another domain will likely be orthogonal to all other tRNA:synthetase pairs in an organism's translation machinery.

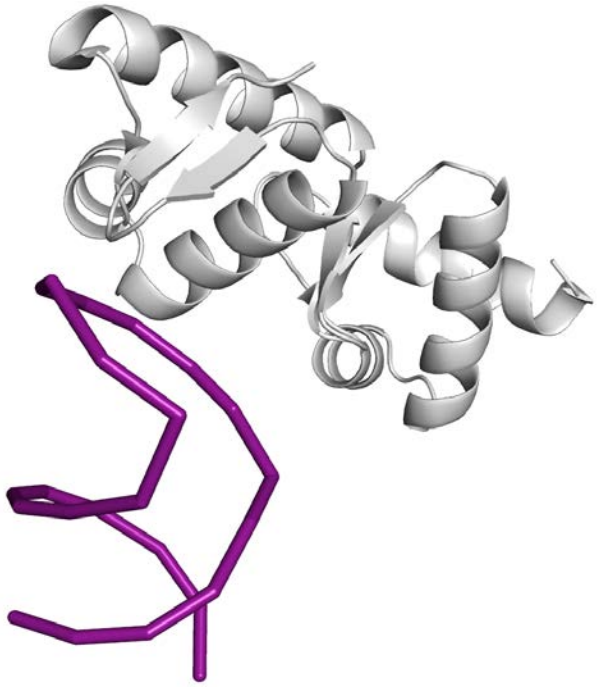

Figure 1.5 SelB crystal structure (gray) complexed with $tRNA^{Sec}$ (purple).

Once an orthogonal pair has been identified, it is engineered specifically for the ncAA of choice. To begin, the tRNA anticodon loop must be altered to a codon that will be exclusively assigned to the ncAA, often the amber stop codon. Additionally, the synthetase must be engineered to exclusively recognize the corresponding tRNA and ncAA of choice.[1] While this method has been used to successfully translate many ncAAs, it is not sufficient for all ncAAs. Even with the addition of an effectively engineered tRNA and synthetase, EF-Tu's proofreading capabilities can still prevent ncAA translation (Figure

1.4).[7, 34] In such cases, the orthogonal tRNA/synthetase pair method must be supplemented by the addition of an EF-Tu which will accommodate the ncAA.[16, 17]

An example in nature of an elongation factor exclusive for one amino acid is SelB (Figure 1.5). *In vivo*, SelB, a paralog of EF-Tu, is specific for the twenty-first amino acid, selenocysteine.[35] EF-Tu does not participate in the translation of selenocysteine. Rather, organisms that can directly translate selenocysteine using the ribosomal translation machinery do so using SelB as an elongation factor exclusive for the delivery of selenocysteinyl-tRNA to the ribosome.

Similarly, EF-Tu must be engineered by researchers to allow certain ncAAs to be efficiently incorporated by the translation machinery. Phosphoserine (Sep) is an example of an ncAA that required an exclusive EF-Tu.[17] An EF-Tu specifically engineered to accommodate the desired ncAA can be used in conjunction with an orthogonal tRNA/synthetase pair, creating an orthogonal triplet specific for the ncAA. Translation of Sep requires an orthogonal triplet, tRNA, synthetase, and EF-Tu, in order to be efficiently translated in an *in vivo* system. To engineer EF-Tu for Sep, random mutagenesis was used to alter the amino acid binding pocket of EF-Tu. An *in vivo* selection process was then used to identify an EF-Tu able to specifically recognize and accommodate Sep. However, this method would require a novel, engineered EF-Tu for each ncAA. Instead, directed evolution has been expanded by incorporating evolutionary information into protein library design.

In the case of EF-Tu, one plan of attack would be to engineer an orthogonal EF-Tu for each ncAA; however, an alternative would be to employ evolutionary data to develop more promiscuous EF-Tus. These EF-Tus would be able to recognize a broader range of

substrates, in this case, a broader range of amino acids, specifically ncAAs. Higher-order directed evolution methods which use ancient evolutionary information have been used to engineer proteins with increased promiscuity.

1.4 Ancestral sequence reconstruction

In addition to orthogonal translation components and directed evolution of proteins, fresh approaches that exploit ancient evolutionary information are proving their utility in protein engineering. Instead of using an exhaustive or random search for mutations that impart novel functionality, researchers can use ancestral sequence information to mine sequence space that has already been tested for viability by nature. These higher-order directed evolution strategies, such as ancestral sequence reconstruction (ASR), can be used to direct protein engineering to design variants with desirable phenotypes such as increased stability or promiscuity. These traits, especially increased promiscuity, may be useful to engineer translation machinery which is better able to accommodate ncAAs.

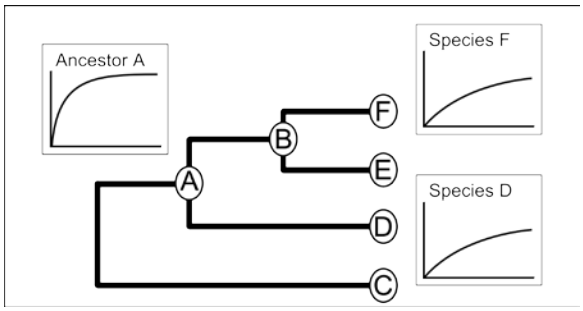

Figure 1.6 Phylogenetic tree showing increased Km and Vmax for ancestor A. This example of a protein phylogeny illustrates an improved Km and Vmax for ancestral proteins over modern proteins. All graphs show the rate of reaction versus substrate concentration. Modern proteins from species F and D show low Km and Vmax. The ancestral protein inferred at node A shows an increase Km and Vmax.

ASR is a computational technique that can rewind the tape of evolution and resurrect ancient proteins that theoretically once existed. This method uses extant protein sequences to computationally determine protein sequences that existed at various speciation events.[36] The advantage of this technique is that these ancestral proteins have already been tested by nature and deemed fit to survive. In several labs, ASR has been successfully used to identify protein sequences with increased thermostability, promiscuous enzymatic activity, or greater stability during directed evolution.[37-39]

One way ASR has facilitated protein engineering is by directing the design of more staple protein variants. For example, ancestral proteins can have increased thermostability and kinetic stability which is thought to be derived from living on a hotter planet during the Precambrian period.[37, 38] Over billions of years, proteins evolved in conditions that were slowly becoming more moderate in terms of temperature, leading to modern proteins which are often stable only within a very narrow temperature range (Figure 1.6).

Additionally, ASR has produced ancestral proteins that demonstrate greater substrate promiscuity. Resurrected β-lactamases from the Precambrian period were found

not only to have resistance to penicillin antibiotics, but they also demonstrated activity against third-generation antibiotics. In contrast, modern β-lactamases have more limited enzymatic capabilities and are not effective against third generation antibiotics.[38] Since promiscuous function in enzymes is thought to be advantageous when seeking to evolve novel protein function, ancestral proteins may be better scaffolds to use in directed evolution protein engineering.[40]

Another technique combines directed evolution and ASR by introducing stabilizing, ancestral mutations into an evolving protein thereby compensating for the destabilizing mutations which seek novel function. Mutations which lead to promiscuous or novel function are generally destabilizing and often reduce or eliminate enzymatic activity. By bringing a protein sequence closer to the ancestral sequence, the protein can better mitigate the effects of destabilizing mutations that arise during directed evolution.[39] This ancestral-modern hybrid can serve as a better scaffold on which to introduce mutations during directed evolution. A more stable protein can accumulate a wider range of mutations thereby allowing a broader search of sequence space, the better to seek novel enzymatic function.

The increased promiscuity of ancestral proteins suggests ancestral protein libraries present profitable sequence space to mine for the ability to translate ncAAs. Although the modern translation machinery has evolved high specificity for only twenty amino acids, ancient translation components may be more promiscuous and may accept a wider range of substrates possibly including ncAAs. Additionally, these ancient sequences can direct stabilizing mutations in engineered proteins which may compensate for destabilizing mutations while seeking novel function via directed evolution.

Methods to incorporate ncAAs continue to evolve as researchers' understanding of protein synthesis continues to develop. Initially the tRNA and synthetase were targeted; researchers knew an orthogonal pair was required for each amino acid regardless of whether that amino acid was natural or unnatural. The concept of adding orthogonal translation components was expanded to include engineered EF-Tus, developed using directed evolution. Currently, higher-order directed evolution offers researchers ancestral protein sequences as a new way to approach ncAA translation. Ancestral sequences can offer both increased stability and promiscuity thereby acting upon a wider range of substrates than modern protein sequences. Stabilizing, ancestral mutations can also compensate for destabilizing mutations, a vital component when seeking altered or novel function via directed evolution. Ultimately, the way directed evolution impacts ncAA translation will continue to develop as scientists' understanding of both translation and evolution improves.

1.5 **Approach**

These advanced protein engineering strategies present the opportunity to alter EF-Tu's substrate specificity. Within the framework of this strategy, there are two approaches to broaden EF-Tu's substrate acceptance. One is to knockout EF-Tu's proofreading capabilities to develop a variant that can accommodate additional ncAAs. This method, however, requires a trade-off between the degree of polyspecificity desired to translate ncAAs and the specificity required for host organism survival. Herein, we present an alternative strategy: engineering a novel EF-Tu with broader ncAA compatibilities to be used in complement with native EF-Tu. This strategy parallels an evolved mechanism for cellular co-translational incorporation of selenocysteine (Sec) which uses a dedicated

elongation factor, SelB, in concert with EF-Tu. This precedent employing SelB supports our approach to engineer a substrate-promiscuous EF-Tu that can work in parallel with native elongation factors.

Computational methods that exploit models of molecular evolution have been previously employed to develop enzymes with expanded substrate scope. These strategies are based on the concept that enzymes evolved specialized activity from generic activities, a theory that is supported by research demonstrating ancestral proteins exhibit broader substrate compatibility than their modern counterparts.[38, 40] To develop polyspecific EF-Tu variants, we employed multiple computational methods predicated on the assumption that ancestral EF-Tu might have demonstrated expanded substrate scope. One method, ancestral sequence reconstruction (ASR), infers ancestral protein sequences based on theories of evolution. This library was developed by Dr. Eric Gaucher.[37] We also used an EF-Tu library developed by Dr. Megan Cole based on reconstructing evolutionary adaptive paths (REAP). The REAP method assumes, based on sequence similarity, that SelB and EF-Tu are paralogues.[41] This in turn suggests EF-Tu and SelB share a common ancestor that exhibited broader substrate-compatibility than modern elongation factors. Given this reasoning, EF-Tu and SelB are suitable protein families for REAP analysis toward engineering a promiscuous EF-Tu. Additional library techniques include consensus sequences developed by Dr. Cole and a binding-site transplantation library developed by myself. These EF-Tu libraries were complemented by two tRNA library based on $tRNA^{Val}$ and $tRNA^{Sep}$. These libraries were generated with the intention of enabling us to examine aa-tRNA:EF-Tu binding from the perspective of both the protein and the tRNA.

Herein, we describe our efforts to transform how EF-Tu is utilized to incorporate ncAAs. Leveraging computational methods, we present data on four EF-Tu libraries and two tRNA libraries used in both *in vitro* and *in vivo* translation. We review methods used for EF-Tu library development (Chapter 2), *in vitro* assay development and optimization (Chapter 3 and 4), and the *in vivo* assay (Chapter 5). Briefly, we touch on further directions taken to continue the project (Chapter 6) and reflect upon possible future directions (Chapter 7).

Ultimately, the *in vivo* assay led to the identification of EF-Tu variants which accept non-native substrates. By mass spectrometry, we demonstrate two variants have expanded substrate acceptance. Fitness tests show a polyspecific EF-Tu is an accepted addition to the translation machinery and improves host organism fitness relative to an ncAA-specific engineered EF-Tu. These results lend credence to our choice of computational method (in this case REAP) and also suggest that EF-Tu and SelB may have a shared, substrate-promiscuous ancestor. Overall, we believe this approach complements current research highlighting the advantages of improved OTSs and promotes a more comprehensive approach critical to achieving future goals.

Portions excerpted from a review article previously published as Cox, V.E. and Gaucher, E.A. 2015. Molecular evolution directs protein translation using ncAAs. *Curr. Protoc. Chem. Biol.* 7:223-238. doi: 10.1002/9780470559277

CHAPTER 2. EF-TU LIBRARY DESIGN: ASR, CONSENSUS, REAP, AND BINDING-SITE TRANSPLANTATION

2.1 Introduction to library design

While several methods have been successfully used to develop ncAA-specific EF-Tu variants, in order to design a substrate-promiscuous EF-Tu a different strategy was required. To explore the mutations required to permit ncAA translation, EF-Tu variants were developed using four methods of protein library design: consensus sequence design, ancestral sequence reconstruction, reconstructing evolutionary adaptive pathways, and binding-site transplantation.[36, 41, 42] These techniques offer a variety of methods all intended to generate small, phenotypically-diverse libraries with high functionality. The consensus sequence, as well as ASR, mutate residues throughout the secondary structure of the protein. The REAP method identifies amino acid residues associated with altered protein function, and the binding-site transplantation library targets the six residues comprising the amino-acid binding site.

Table 2.1 Table of computational methods used to develop EF-Tu libraries. Libraries are listed along the left side; library characteristics are listed across the top.

Strategy	affects entire protein	targets amino acid binding site	considers ancestral sequence	considers modern sequence	possible increased promiscuity
Ancestral sequence reconstruction (ASR)	X		X		X
Reconstructing Evolutionary Adaptive Paths (REAP)		X	X	X	X
Consensus	X			X	
binding-site transplantation		X		X	

Each of these methods has different benefits and drawbacks (Table 2.1). Consensus sequence construction is compatible with a low-throughput assay, but it only samples

extant protein sequences and can be negatively impacted by ignoring epistatic interactions. It also results in only a single variant; however, that variant can be a viable scaffold for future mutations. Conversely, ASR computationally infers ancestral sequences and can be used to develop a larger library. Previous work has found that ancestral sequences can have greater thermostability and may exhibit greater polyspecificity; however, these variants might not be compatible with modern translation factors (e.g. tRNAs, ribosome, and other elongation factors).[38, 41] REAP is the only technique that samples both modern and ancestral sequence space. This method focuses on generating variants with diverse phenotypes by targeting sites of functional divergence. However, in this case REAP only infers possible mutation sites; it does not recommend what amino acids should be used as substitutions. To limit the library size, we used alanine scanning, but if using site-saturation or many combinations of mutations, a REAP library can become too large for low-throughput screening techniques. The binding-site transplantation library explores the possibility that promiscuous amino acid binding sites may be naturally occurring in other species. This library replaces the six residues comprising the amino acid binding pocket in E. coli EF-Tu, with binding site residues from EF-Tus found in other species to examine if other species have more promiscuous binding pockets.

2.2 Consensus

Generating a consensus sequence involves sampling homologous extant sequences of a protein to generate a single sequence featuring the most common amino acid at each position in the protein. A significant advantage of using a consensus sequence is that it can be highly stable and thus can make exceptionally good scaffolds on which to make further potentially destabilizing mutations.[36] This sequence was developed by Dr. Cole.

2.3 ASR

ASR is a computational methodology used to infer ancestral proteins using extant, (modern) protein sequences as input. There is evidence to suggest that ancestral proteins can offer a number of advantages (e.g. thermostability).[37] In particular, some ancestral proteins have been found to accept a broader range of substrates.[38] Based on these data, it is reasonable to theorize that ancestral EF-Tu variants may bind a broader range of substrates, specifically ncAAs. Our computational process and library design are described elsewhere.[37, 41] This library was developed by Dr. Gaucher.

2.4 REAP

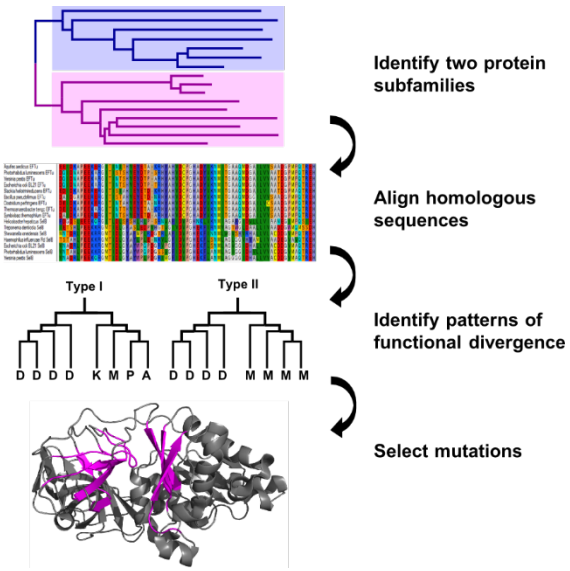

Figure 2.1 Overview of REAP method. A phylogeny of two homologous protein families is shown. An MSA is generated and sites of functional divergence are identified via DIVERGE software and mutated to generate protein variants.

REAP has been previously used to guide development of enzyme libraries with expanded substrate-acceptance.[43] In brief, this method uses inferred amino acid mutation

rates to predict which amino acid substitutions are most likely to impart novel protein activity (

Figure 2.1). REAP analysis is based on two complementary assumptions. First, that amino acids which impact function are conserved during the evolution of a protein family. Second, residues lacking conservation are likely not relevant to activity. REAP functions by ranking residues according to their degree of conservation. Amino acid sites with low inferred mutation rates are predicted to have a high correlation to function and thus should be targeted during library design. Correspondingly, sites with high mutation rates are predicted to minimally contribute to new protein behaviors and can be excluded from library design. As a result, inferred mutation rates allow for the design of a targeted library enriched with functional variants. A central tenet of this method is that a REAP-developed library can retain the functional diversity of a larger library, while minimizing the number of variants that will not exhibit new function. This method can lead to the design of a small, phenotypically-diverse protein library with a large proportion of active variants.

Conserved amino acid sites are classified during REAP analysis as exhibiting either Type I or Type II functional divergence. Type I indicates an amino acid is conserved for only one branch of a protein phylogeny.[44, 45] This indicates the residue is critical for function in one protein family (where it is conserved), but not the other (in which the site is variable). Alternatively, amino acid sites exhibiting Type II functional divergence show conservation in both branches of the phylogeny, but the amino acid identity differs between families.[46] This type of divergence suggests that while the amino acid position is important to protein activity in both families, its role in protein behavior differs.

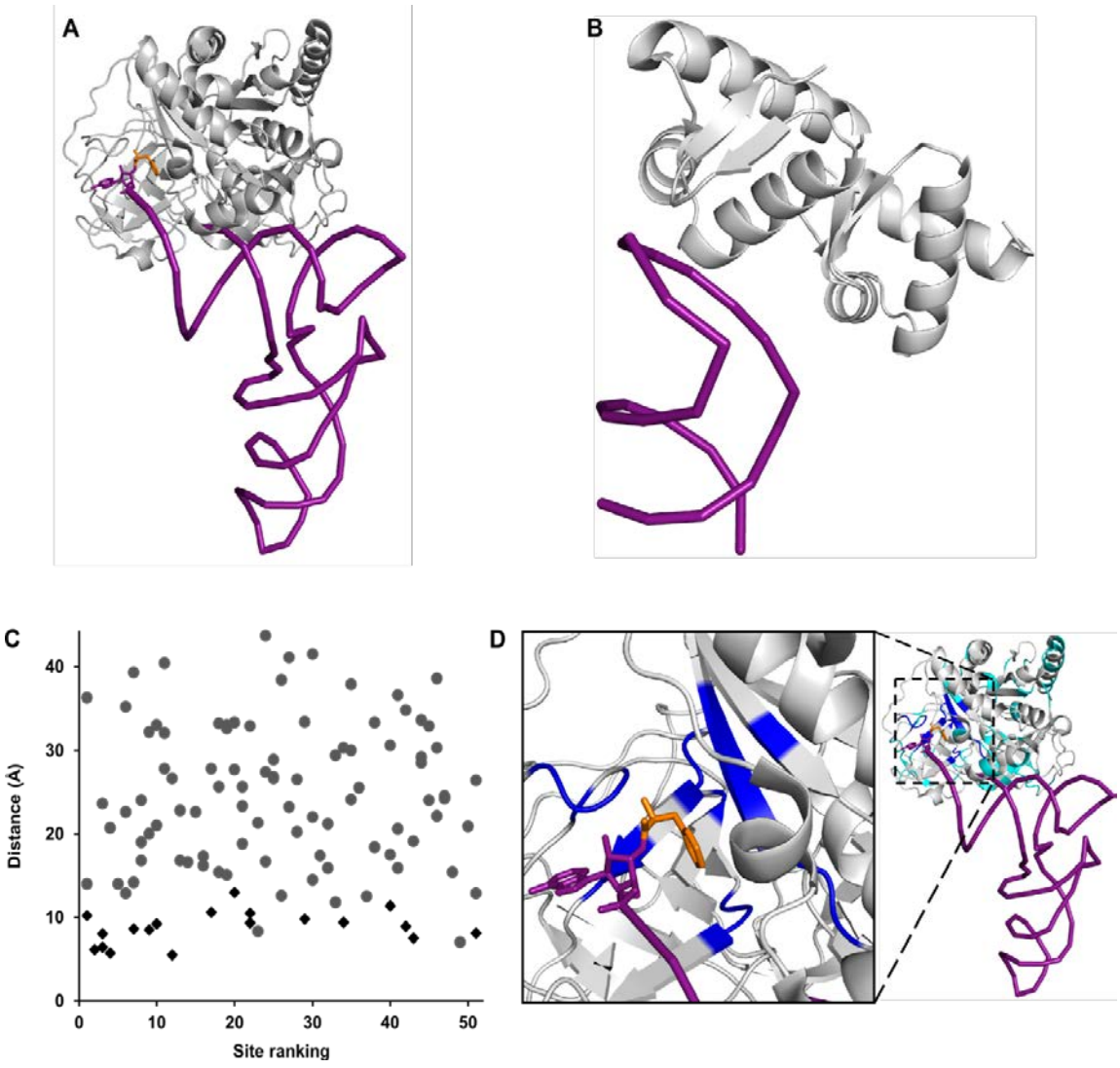

Figure 2.2 Stepwise development of the REAP-designed EF-Tu library. (A) Crystal structure of EF-Tu (gray) complexed with $tRNA^{Phe}$ (purple). Phenylalanine (orange) situated in amino acid binding pocket. Based on Protein Data Bank structure 1OB2. (B) Crystal structure of SelB (gray) complexed with $tRNA^{Sec}$ (purple). Based on Protein Data Bank structure 2PJP. (C) Plot shows residues identified by REAP. Ranking of position versus distance from target. Black diamonds denote positions selected for mutation. Gray circles indicate amino acid sites that were not included in protein library. For more detailed information, refer to table. (D) Amino acids identified by REAP analysis. Sites mutated to generate EF-Tu library (blue). Residues not selected for library (cyan). EF-Tu:$tRNA^{Phe}$ color scheme the same as in panel A.

Dr. Cole designed the REAP library by comparing EF-Tu and SelB sequences. Examination of sequence similarity suggests that EF-Tu and SelB can be classified as functionally divergent homologues making them appropriate protein families for this

application. EF-Tu and SelB sequences from 19 prokaryotic families were aligned and evaluated to identify amino acid positions predicted to influence substrate scope. The aligned sequences, along with phylogenies of each protein subfamily, were entered into DIVERGE software for analysis via three computational models (Figure 2.3, Figure 2.4).[47] Two models, Gu99 and Gu01, identify Type I functional divergence, in each case employing different parameters for analysis. Sites associated with Type II functional divergence were identified by a third model. Residues were ranked according to their posterior probability (Type I) or posterior ratio (Type II). This produced a rank-ordered list of residues with the top members being predicted to have a greater influence on activity (

Table 2.2).

A preliminary list of residues for mutation was produced by parsing the top-ranked 112 residues according to their distance from the target. Distances were calculated using the Cγ of the binding target amino acid and the Cα of the EF-Tu residue. Residues exceeding 13Å were removed from the list leaving 26 predicted positions in close proximity to the target. Of the 26 sites, 7 residues were excluded thereby culling the final list to 19 residues. Residues rejected included aliphatic residues between 12-13Å since they were not expected to have a significant effect on substrate acceptance. Methionine residues were also discounted due to their integral role in protein structure.[48] Lastly, although the Cα of Y76 was within 13Å, the Cγ of the side chain fell outside the distance cutoff suggesting that mutation of the residue would not impact the structure of the amino acid binding pocket (Figure 2.2).

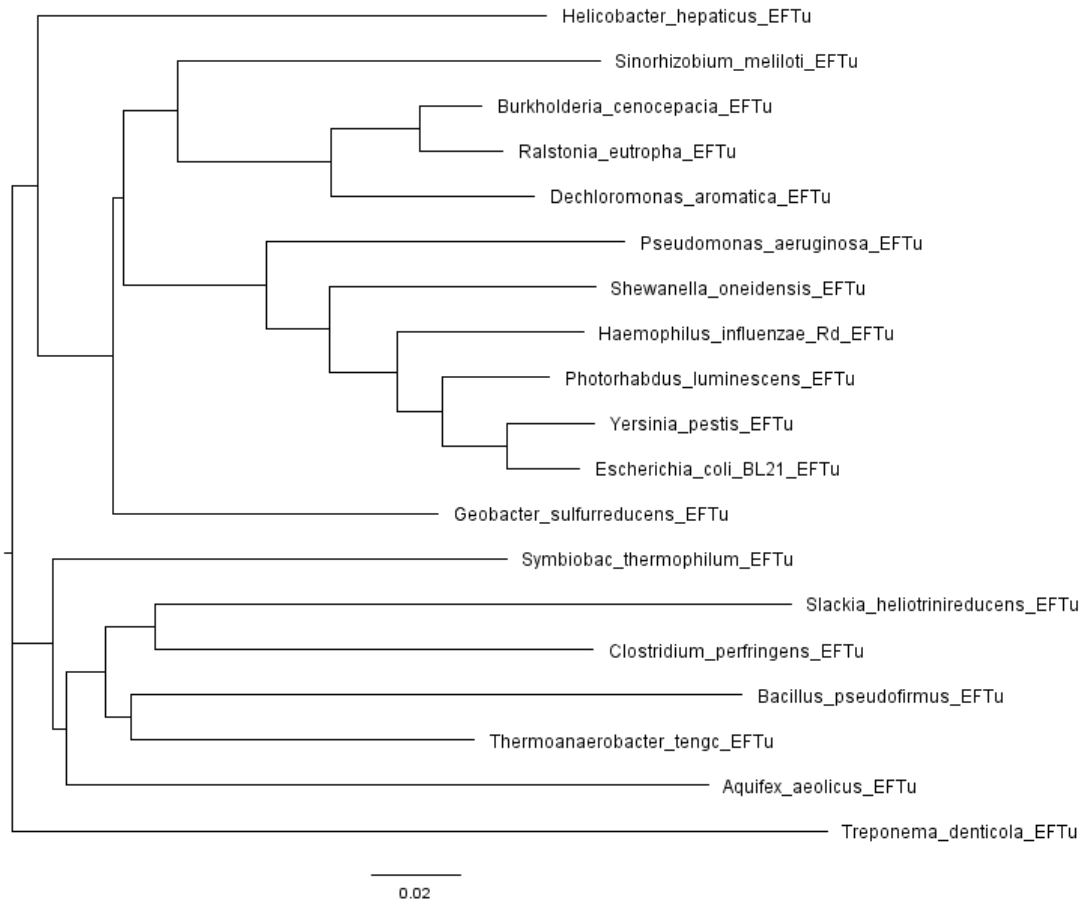

Figure 2.3 Phylogeny used for REAP analysis based on EF-Tu sequences.

In addition to the 19 residues selected via REAP, positions N13, V227, and V274 were also included in library design. Since protein lengths were normalized during computational analysis, the first 47 residues were appraised by eye and site N13 was chosen for exhibiting Type I functional divergence. Sites V227 and V274 were selected based on proximity to the target, thereby bringing the total to 22 mutation sites targeted in the REAP library (Figure 2.5).

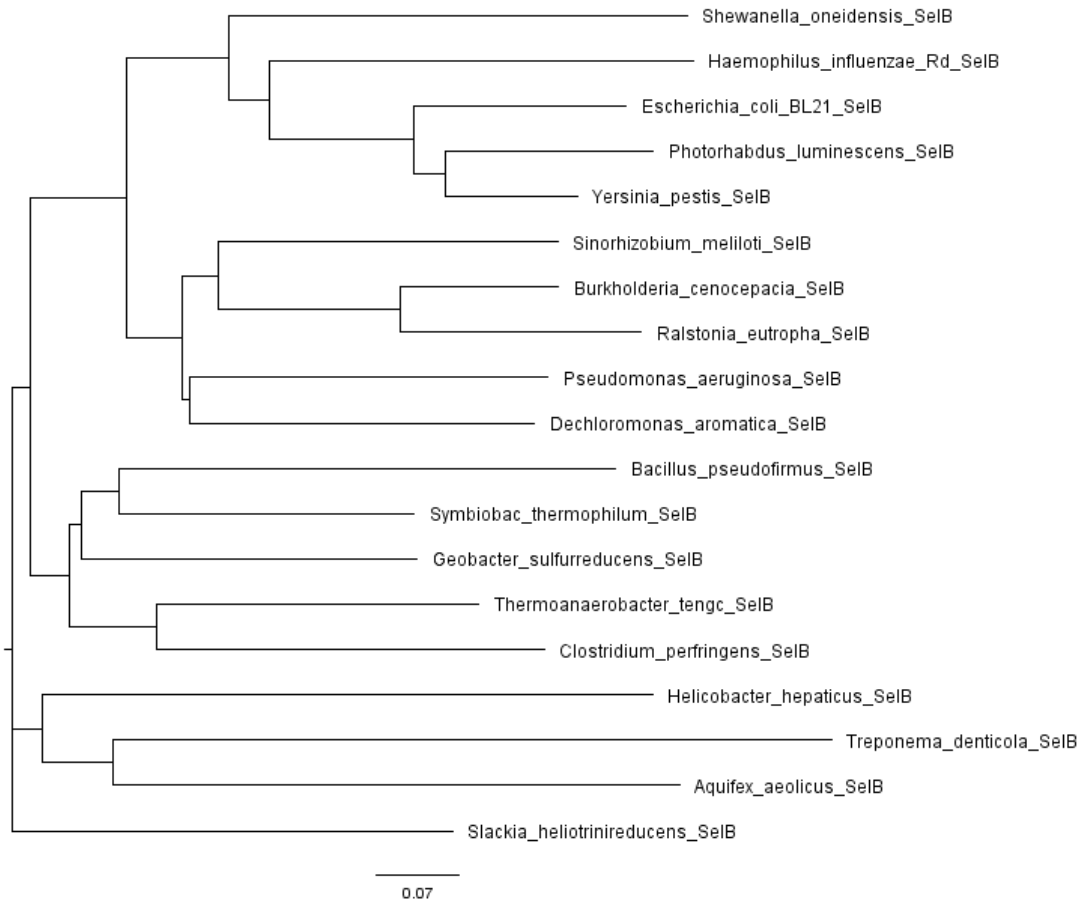

Figure 2.4 Phylogeny used for REAP analysis based on SelB sequences.

Table 2.2 Top-ranked REAP-identified positions. Amino acid numbers relative to position in E. coli sequence. Yellow highlight indicates residues mutated in library.

Rank	Type I (Gu99)	Type I (Gu01)	Type II
1	276	359	81
2	359	276	273
3	189	65	275
4	65	240	261
5	240	189	93
6	367	305	277
7	342	48	259
8	202	255	85
9	154	254	216
10	56	304	230
11	142	156	106
12	364	200	66
13	251	342	211
14	304	142	256
15	200	154	254
16	236	367	333
17	338	202	64
18	96	389	372
19	232	361	108
20	150	63	172
21	279	105	174
22	263	303	78
23	196	106	77
24	344	176	290
25	385	296	329
26	148	338	49
27	144	236	86
28	239	144	135
29	159	263	215
30	323	51	331
31	296	148	233
32	305	122	223
33	76	385	363
34	379	56	272
35	163	147	124
36	330	150	51
37	254	108	268
38	327	196	71
39	255	327	389
40	100	214	111
41	326	266	332
42	149	149	68
43	67	127	342
44	378	306	301
45	48	295	391
46	328	141	130
47	147	374	57
48	63	287	328
49	266	260	378
50	214	199	326
51	299	220	262

amino acid position	13	63	64	65	66	67	68	78	214	215	216	227	230	259	261	262	263	272	273	274	275	276
EF-Tu																						
Escherichia coli BL21	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	V	G	V
Aquifex aeolicus	N	N	I	T	H	V	E	H	I	E	D	V	R	E	F	R	K	D	N	I	G	V
Bacillus pseudofirmus	N	S	T	A	H	V	E	H	V	E	D	A	R	E	F	R	K	D	N	I	G	A
Burkholderia cenocepacia	N	N	T	A	H	V	E	H	V	E	D	V	R	E	F	R	K	D	N	V	G	I
Clostridium perfringens	N	N	T	A	H	V	E	H	V	E	D	A	R	E	F	R	K	D	N	I	G	A
Dechloromonas aromatica	N	N	T	A	H	V	E	H	V	E	D	V	R	E	F	R	K	D	N	I	G	A
Geobacter sulfurreducens	N	A	T	S	H	V	E	H	V	E	D	A	R	E	F	R	K	D	N	I	G	A
Haemophilus influenza Rd	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	I	G	A
Helicobacter hepaticus	N	A	T	S	H	I	E	H	V	E	D	V	R	E	F	R	K	D	N	V	G	I
Photorhabdus luminescens	N	S	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	V	G	V
Pseudomonas aeruginosa	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	V	G	I
Ralstonia eutropha	N	N	T	A	H	V	E	H	V	E	D	V	R	E	F	R	K	D	N	V	G	L
Shewanella oneidensis	N	N	T	S	H	I	E	H	I	E	D	V	R	E	F	R	K	E	N	C	G	I
Sinorhizobium meliloti	N	S	T	A	H	V	E	H	I	E	D	V	R	E	F	R	K	D	N	I	G	A
Slackia heliotrinireducens	N	S	I	A	H	I	E	H	V	E	D	A	R	E	F	R	K	D	N	I	G	C
Symbiobac thermophilum	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	D	N	V	G	A
Thermoanaerobacter tengc	N	N	T	T	H	V	E	H	V	E	D	A	R	E	F	R	K	D	N	I	G	V
Treponema denticola	N	N	T	R	H	L	E	H	I	E	D	V	R	E	F	N	K	D	N	V	G	L
Yersinia pestis	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	V	G	V
SelB																						
Escherichia coli BL21	I	D	L	G	Y	A	Y	F	I	D	R	V	T	H	Q	N	Q	Q	R	I	A	L
Aquifex aeolicus	L	D	I	G	F	A	Y	I	V	D	S	L	S	Q	H	G	V	E	R	I	A	L
Bacillus pseudofirmus	T	E	L	G	Y	A	P	I	I	D	Q	V	T	Q	H	H	Q	Q	R	V	A	I
Burkholderia cenocepacia	I	E	L	G	Y	A	Y	L	V	D	R	V	T	H	Q	N	R	E	R	C	A	L
Clostridium perfringens	I	N	L	G	F	T	F	I	V	D	R	A	T	Q	H	D	E	Q	R	C	A	L
Dechloromonas aromatica	I	D	L	G	Y	A	Y	F	V	D	R	V	T	R	Q	D	A	Q	R	I	A	L
Geobacter sulfurreducens	I	E	L	G	F	A	H	I	V	D	R	V	T	Q	H	G	R	Q	R	V	A	V
Haemophilus influenza Rd	I	D	L	G	Y	A	Y	F	I	D	R	V	T	H	Q	N	T	Q	R	L	A	L
Helicobacter hepaticus	I	D	L	S	F	S	H	F	I	D	R	V	S	Q	H	D	K	H	R	V	A	L
Photorhabdus luminescens	I	D	L	G	Y	A	Y	F	I	D	R	V	T	H	Q	N	Q	Q	R	I	A	L
Pseudomonas aeruginosa	I	D	L	G	Y	L	Y	F	V	D	R	V	T	H	Q	N	R	Q	R	V	A	L
Ralstonia eutropha	I	E	L	G	Y	A	Y	L	V	D	R	V	T	H	Q	N	R	Q	R	C	A	L
Shewanella oneidensis	I	D	L	G	Y	A	F	F	L	D	R	V	T	H	Q	G	M	T	R	V	A	L
Sinorhizobium meliloti	I	D	L	G	F	A	Y	F	V	D	R	V	T	H	Q	N	Q	Q	R	C	A	L
Slackia heliotrinireducens	V	E	L	G	F	A	R	V	V	D	R	I	T	Q	H	D	K	N	R	V	A	L
Symbiobac thermophilum	I	D	I	G	F	A	R	V	I	D	R	V	T	Q	H	G	E	Q	R	V	A	V
Thermoanaerobacter tengc	I	D	L	G	F	A	Y	I	V	D	R	V	T	Q	H	E	R	Q	R	T	A	I
Treponema denticola	I	E	L	G	F	A	S	I	V	D	R	V	T	Q	H	H	K	T	R	T	A	L
Yersinia pestis	I	D	L	G	Y	A	Y	F	I	D	R	V	T	H	Q	N	Q	Q	R	I	A	L

Figure 2.5 Multiple sequence alignment of EF-Tu and SelB substrate binding site. Type I (green), Type II (orange) and residues chosen for proximity (navy) are shown

This selection of 22 residues represents a larger targeted sequence space than previous efforts. However, despite scanning more residues for activity, mutations were combined to generate a small, targeted library. REAP variants were built on a wild type *Escherichia coli* EF-Tu backbone and each contained 4, 8, or 12 mutations (Figure 2.6). This approach contrasts with previous publications which have typically targeted a more limited number of amino acid sites, but in greater combinations (e.g., via site-saturation) leading to much larger libraries.[11, 17] In line with prior research, the library was developed using alanine scanning to expand the physical space within the amino acid binding pocket.[6,

[11, 16] Since increased flexibility in the binding pocket can broaden substrate acceptance, glycine could have been a promising amino acid replacement instead of alanine.[6, 49] However, because each variant was designed with multiple mutations, replacing residues with glycine was predicted to have a detrimental effect on protein structure.

A

variant	amino acid position																					
	13	63	64	65	66	67	68	78	214	215	216	227	230	259	261	262	263	272	273	274	275	276
E. coli	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	V	G	V
EF-4A		A									A						A		A			
EF-4B			A					A							A			A				
EF-4C				A						A		A										A
EF-4D	A		A		A									A								
EF-4E						A			A				A								A	
EF-4F							A	A								A			A			
EF-8A	A			A		A				A			A		A	A			A			
EF-12A		A	A		A			A	A		A	A		A			A	A		A		A

Figure 2.6 Mutations made to generate REAP library. Protein variants are named by the number of mutations (four, eight, or twelve) followed by letters to distinguish among variants with four mutations. Sequence of wild type *E. coli* EF-Tu is shown for reference.

2.5 Binding-site library development

The binding-site transplantation library was developed by mutating the six amino acid residues in EF-Tu's amino-acid binding site. Previous research has demonstrated the potential to engineer an EF-Tu variant specific for a particular ncAA by mutating only these six residues.[17] Based on this research, we hypothesize that if mutating these six residues can make EF-Tu more specific, different mutations in the same positions might render the protein polyspecific with broader substrate specificity. To test this hypothesis, we created a new EF-Tu variant library which reflects amino-acid binding sites found in EF-Tu proteins from different species.

Residue number	66	215	216	218	228	273
E. coli / bacteria consensus	H	E	D	F	T	N
Nanoarchaeum equitans (1)	H	Q	S	L	T	N
Archaea consensus (2)	H	Q	D	Y	V	N
Nanoarchaeota (3)	A	G	D	Y	T	N
Korarchaeum cryptofilum (4)	H	Q	D	F	V	N
Thermofilum and *Sulfolobus* (5)	F	Q	D	Y	V	N
Desulfurococcus and *Thermosphaera* (6)	Y	Q	D	Y	V	N
Acidilobus (7)	F	Q	G	L	T	N
Caldisphaera lagunensis (8)	F	Q	N	Y	V	N
Sulfolobus tokodaii (9)	F	Q	E	Y	V	N
Eukaryotes consensus (10)	L	Q	D	Y	V	N
Giardia lamblia (11)	L	Q	D	Y	A	N
Entamoeba histolytica (12)	L	Q	D	Y	C	N
Lactobacillales and Oomycota (13)	H	E	D	F	S	N
Dehalococcoides ethenogenes (14)	H	E	D	F	T	A

Figure 2.7 Binding-site transplantation library. Left column indicates from where each sequence was derived. Amino acid substations in the binding site are color-coded: positively charged (purple), negatively charged (blue), nonpolar (green), and polar (yellow).

Wild type *E. coli* EF-Tu was the starting point for all variants. Using BLAST, we identified 247 unique EF-Tu sequences representing species spanning all three domains of life. These sequences were then aligned using Clustal Omega and the binding sites were compared. These binding sites served as guidance when building the library. For each variant, the *E. coli* amino-acid binding site was mutated to reflect a unique binding site found in a different species, the theory being that perhaps a different species has a more promiscuous EF-Tu capable of translating ncAAs (Figure 2.7). These variants with "transplanted" binding sites were then assayed and characterized. The final library represented nine binding sites from the domain archaea, three from eukaryotes, and three from bacteria including the full sequence of E. coli. These EF-Tu were assayed *in vivo* for compatibility with Sep. Since E. coli EF-Tu is known to be incompatible with Sep, it was

used as a negative control. The relative ability of these variants to translate a gene containing Sep (at an amber stop codon) was compared to the ability of EF-Sep to incorporate Sep.

CHAPTER 3. OPTIMIZING AMINOACYLATION PROTOCOL AND TRANSLATION PROTOCOL FOR TRNA VARIANTS

Herein we discuss optimization of the aminoacylation protocol and translation protocol specifically for work with tRNA variants. This data relates to but does not repeat data presented in other chapters. The *in vitro* translation experiments were performed with a translation kit that lacked tRNAs and amino acids which is different from the kit used in the subsequent chapter which lacked EF-Tus. Because these *in vitro* translation kits contain both cell lysate and recombinantly expressed components, we would not necessarily expect kits lacking tRNA and kits lacking EF-Tu to be optimized the same way.

3.1 Initiated protocol to use flexizymes in lab

Flexizymes are enzymatic RNA sequences that can replace synthetases. However, whereas synthetases recognize their cognate amino acid, flexizymes recognize a leaving group attached to the amino acid.[50] The flexizyme charges the amino acid to the tRNA and the aa-tRNAs can then be used in an *in vitro* translation assay. We determined amino acids with the appropriate leaving group would be ordered from a commercial source. We also ordered the primers and were able to successfully synthesize the flexizymes. However, it became clear the tRNA protocol was not working. The ratio of signal to noise was low with high variability from day to day. Without a functional tRNA assay, the flexizyme work was significantly less applicable to the project. As such, we suspended work on flexizymes to optimize the tRNA aminoacylation and the corresponding *in vitro* translation assay.

3.2 Optimizing the aminoacylation and *in vitro* translation protocols for tRNA variants

3.2.1 Experimental protocols

For these experiments, we used an aminoacylation protocol followed by *in vitro* translation. In the acylation protocol, we used all twenty tRNAs but only one amino acid and one synthetase; thus, we could assume the product contained only one aa-tRNA. Aminoacylations were performed with either radiolabeled S-35 fMet or Val as the target amino acid. Aminoacylations using fMet could be quantified (via scintillation) before the acylation was carried through the translation reaction. With Val, the aminoacylation itself could not be measured; the acylation had to be carried through translation before we could generate any data. With either amino acid, *in vitro* translation was measured by scintillation using radiolabeled S-35 fMet to initiate the peptide. The goal was to assay $tRNA^{Val}$ variants with valine analogs (Figure 3.1).

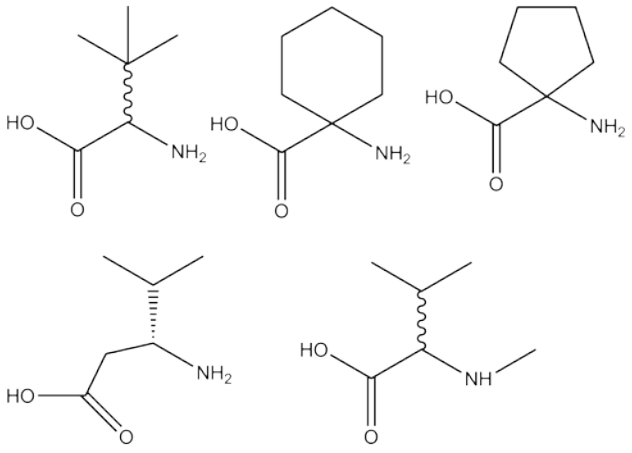

Figure 3.1 Valine analogs that present a variety of challenges to the translation machinery.

In vitro translation was accomplished using a custom PURE translation kit (NEB) from which tRNAs and amino acids had been withheld. To perform *in vitro* translation, we added the aminoacylation reaction (a resolubilized aa-tRNA pellet) and all additional amino acids required to translate the peptide of choice, generally the 11-mer MVEVRH$_6$. Whatever target amino acid was added to the aminoacylation (Val or fMet), this amino acid was not added separately to the *in vitro* translation mix. As such, the only source of the target amino acid during translation would be successfully acylated tRNAs from the aminoacylation reaction.

We always ran negative controls in both the aminoacylation and *in vitro* translation reactions. Negative controls in the aminoacylation protocol lacked the cognate synthetase; either a non-cognate synthetase or no synthetase was added. Thus, negative controls lacked successfully acylated tRNAs. *In vitro* translation could have any of a number of negative controls. The negative control from the aminoacylation could be carried through as a negative control for *in vitro* translation. Additional negative control reactions could lack either the DNA template, one essential amino acid, or all amino acids.

3.2.2 Key accomplishments and challenges

During this assay, we improved the signal to noise ratio by an order of magnitude. Improvements to the protocol included deacylating purchased tRNAs, washing and drying the aa-tRNA pellet, optimizing the volume and solvent used to resolubilize the aa-tRNA pellet, adjusting the ratio of radiolabeled fMet and aa-tRNA storage. Unfortunately, having so many reagents in the aminoacylation and translation protocols meant it was difficult, expensive, and time-consuming to identify and correct sources of low-signal, high background, and data inconsistencies. Because our goal was to use this assay in concert

with our EF-Tu libraries, challenges quantifying our purified EF-Tu variants eventually made this protocol less immediately relevant and we moved on to an *in vivo* assay.

3.2.3 Data

3.2.3.1 Failed controls

When we began work on the *in vitro* tRNA assay, the main problem was the ratio of signal to noise which was roughly one to one (Figure 3.2). Aminoacylation and *in vitro* translation experiments were always run with negative controls.

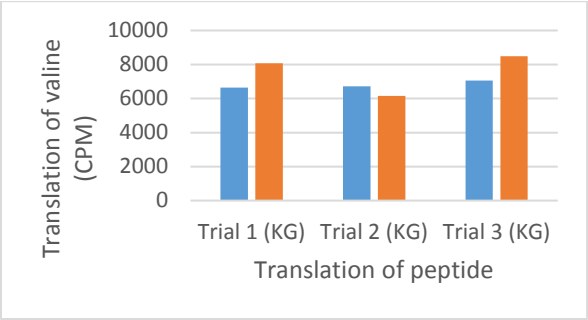

Figure 3.2 Translation prior to optimization. Positive controls (blue bars) and negative controls (orange bars) are shown. Data generated by KG.

The problem was that withholding the synthetase in the acylation reaction ultimately resulted in rates of translation similar to our samples which contained successfully acylated tRNAs (data generated by Kelsey Gratton) (Figure 3.3).

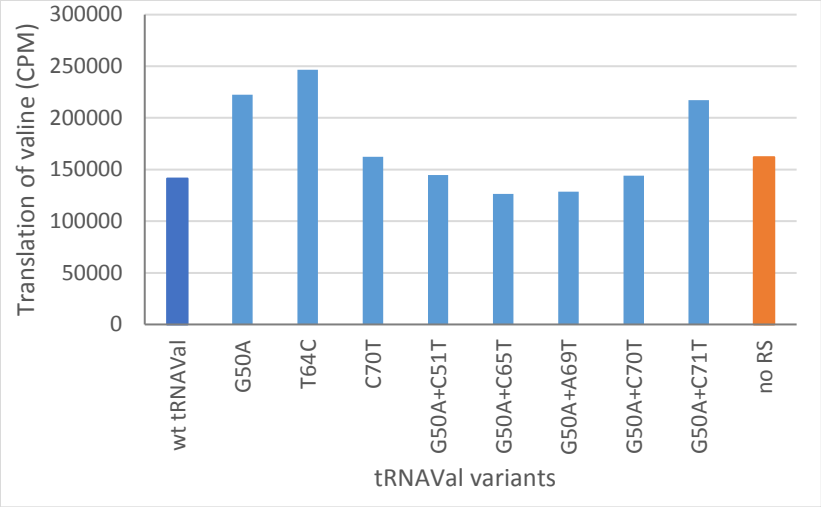

Figure 3.3 Translation of MEVERH$_6$ plasmid with tRNA^{Val} variants. Graph shows a positive control (navy), translation with tRNA^{Val} variants (blue), and a negative control (orange) (data by KG).

3.2.3.2 Protocol optimization

One possible reason for high background was that the tRNA used in the acylation protocol could have already been acylated. Thus, we started to optimize the protocol by deacylating purchased tRNAs and comparing results to tRNAs that were not deacylated. Both samples of tRNA, deacylated and untreated, were processed through the acylation and translation reactions. It appeared the deacylated tRNA gave better data and we decided to proceed using that method (Figure 3.4). However, the background during translation was still sufficiently high to obscure meaning (Figure 3.5).

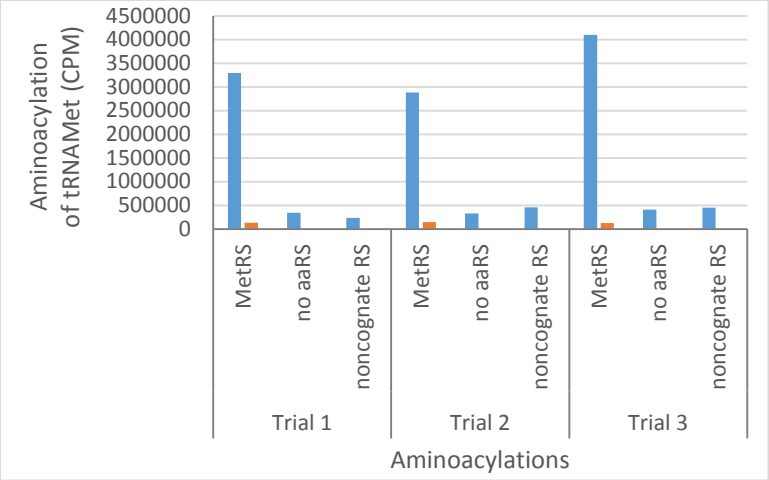

Figure 3.4 Deacylated tRNAs (blue bars) compared to purchased tRNA (orange bars). Graph shows amino acylation of deacylated tRNA with the cognate aaRS, with no aaRS added, and with a non-cognate aaRS (ValRS).

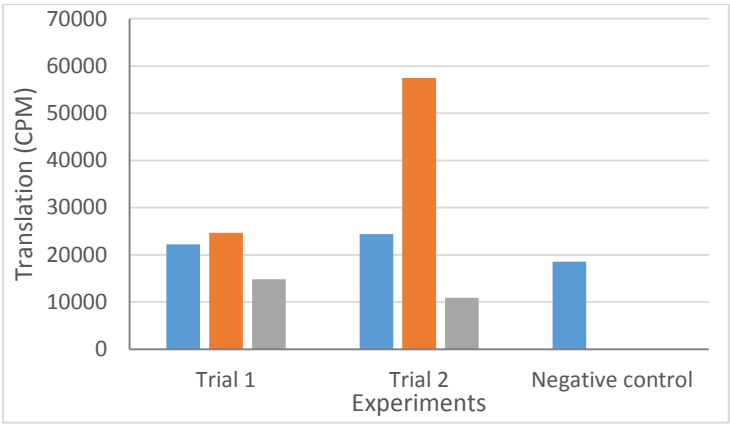

Figure 3.5 Inconsistent valine translation. Translation with the correct ValRS (blue) and with the non-canonical aaRS (orange) and no aaRS (gray). Negative control has no aaRS or tRNA added.

Moving forward, we decided to test a few possibilities at once. In the protocol, after acylating the tRNA with radiolabeled S-35 fMet, the solution was alcohol precipitated, pelleted down, and resolubilized in sodium acetate. We thought there might be residual radiation on the surface of the pellet which might be contaminating the translation mixture

and increasing our background radiation. Residual radiation would make it appear as if we were getting translation with our negative control, but in reality, it might be contamination. To reduce contaminating radiation, we tried washing the pellet twice with 70% ethanol. We also speculated that the sodium acetate used to resolubilize the tRNA could be inhibiting the translation reactions. Perhaps we were not actually getting translation; perhaps the problem was not that background was too high, but rather that our signal was too low. This could account for our poor signal to noise ratio. As such, we also decided to try resolubilizing the pellet in water (Figure 3.6).

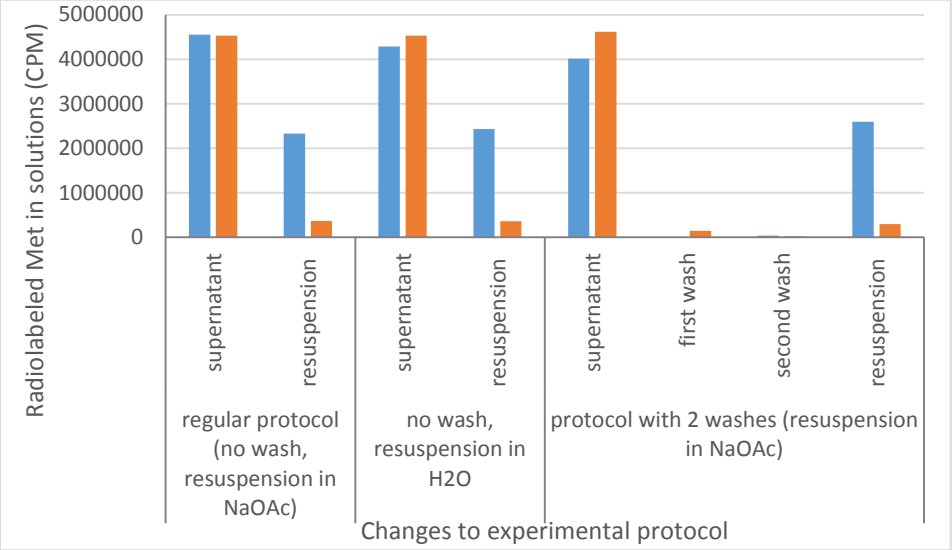

Figure 3.6 Optimizing acylation protocol. Washing pelleted acylation and resolubilizing in sodium acetate (NaOAc) or water with cognate MetRS (blue bars) and with no RS (orange bars).

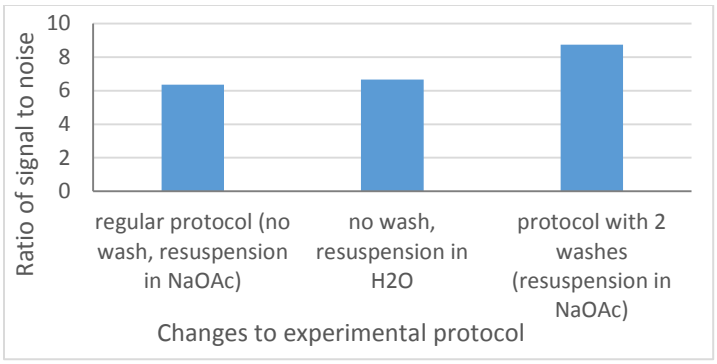

Figure 3.7 Optimizing acylation of Met-RNA. Ratio of supernatant from reaction with MetRS versus supernatant from reaction with no RS.

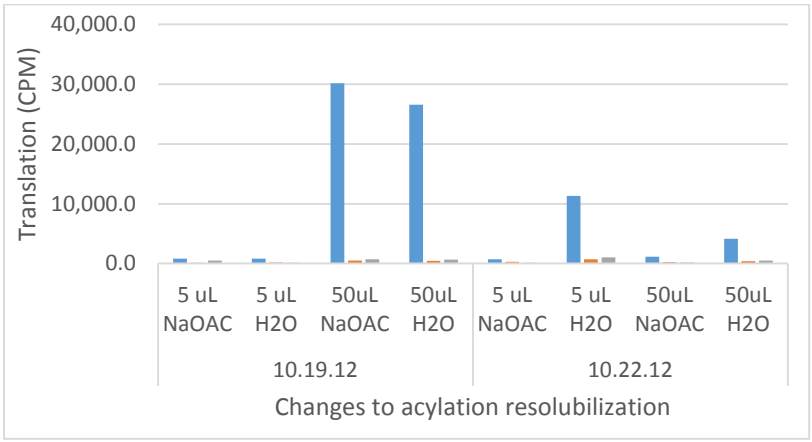

Figure 3.8 Translation Met results. Resolubilizing pellet in 5 or 50 µL of NaOAc or water. Cognate RS, MetRS, (blue). Non-cognate RS, ValRS, (orange). No RS added (gray).

Some improvement was observed in the ratio of signal to noise (ratio of aminoacylation with aaRS versus no aaRS) (Figure 3.7). The data suggested that washing the pellet twice with 70% ethanol did help reduce background radiation. We also decided to try resolubilizing with different volumes. We were originally resolubilizing in 5µL of 1mM NaOAc pH 5.2. We used a small volume to keep the aa-tRNA concentrated but perhaps the small volume did not sufficiently resolubilize the tRNA pellet. We decided to

try using 50µL as a resolubilizing volume. We replicated the translation experiments comparing solvents (Figure 3.8).

Additionally, we experimented using only radioactive fMet as compared to using both radioactive and cold Met together (Figure 3.9).

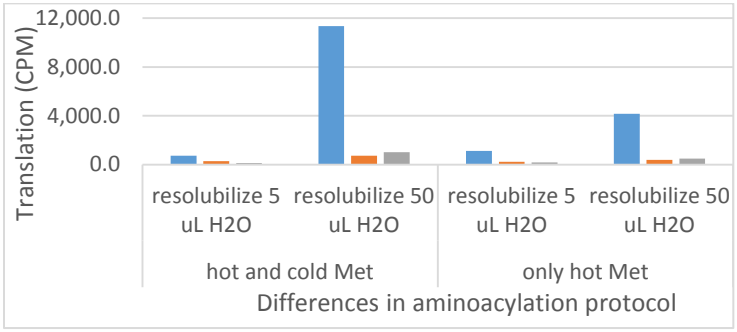

Figure 3.9 Met translation with hot and cold Met and just hot Met. Also tested with various resolubilization volumes. Translation reactions with cognate MetRS (blue), non-cognate ValRS (orange), and no RS (gray).

Following these experiments, we decided resolubilizing in 50µL gave significantly higher signal and also a better signal to noise ratio. The difference between using water and sodium acetate to resolubilize was less clear, but ultimately, it seemed that water was better. We decided to proceed using a combination of radioactive and cold Met which supported previous experiments that suggested using a mixture of hot and cold Met produced the best results. However, the results were still inconsistent.

We decided to try changing the translation protocol to eliminate any possible contaminating amino acids or tRNAs in the NEB translation kit which might affect our results. The new protocol involved starting the translation reaction without our aa-tRNA or radioactive Met. Then, after one hour of translation, we interrupted the incubation and added our aminoacylated tRNA, all nineteen other tRNAs, and radioactive Met. We then allowed translation to continue for three more hours. Using this protocol, contaminating

tRNAs or amino acids would be used up during the first hour of translation and peptides made during this time would not be radioactively labeled. Since radioactive Met was added later, only peptides made with our aminoacylated tRNA would be radioactively labeled. Note, these experiments were each run with four negative controls (Figure 3.10, Figure 3.11).

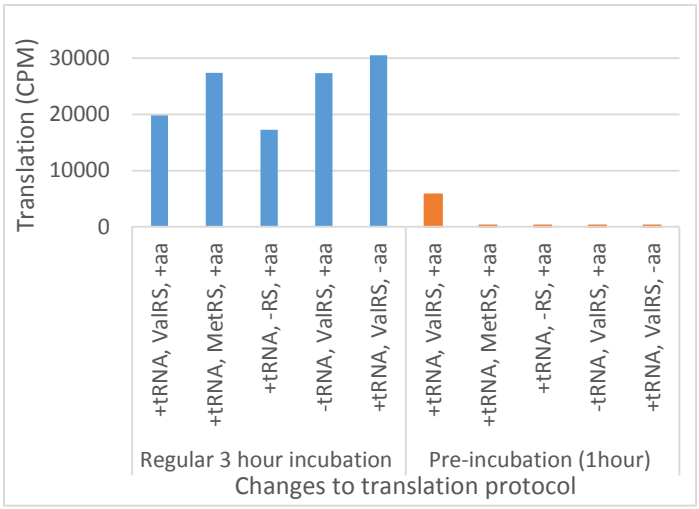

Figure 3.10 Val translation. Val concentration the same as Met (172.9 µM). Translation reactions either incubated for 3-hours or translation mix had an interrupted incubation.

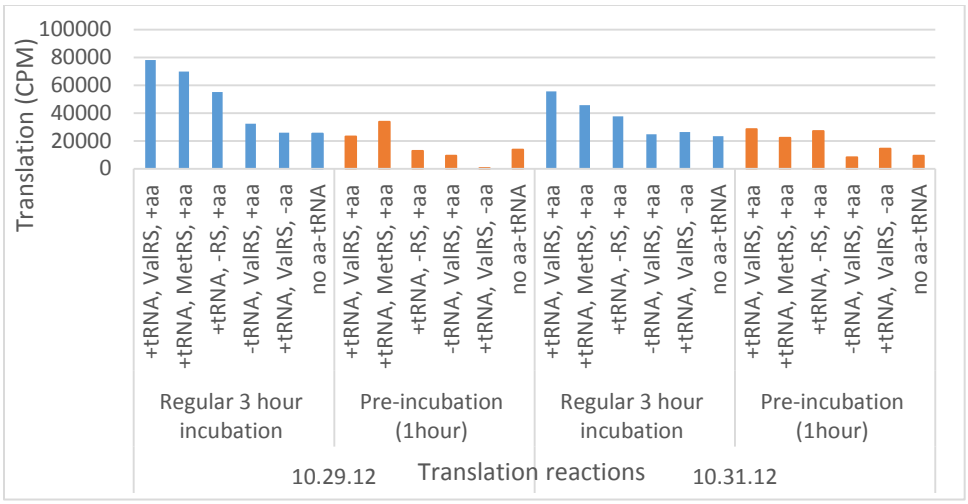

Figure 3.11 Val translation. Translation reactions either incubated for 3-hours or translation mix had an interrupted incubation.

We found that both signal and noise decreased using an interrupted translation protocol, improving background but not the ratio of signal to noise. However, the amount of peptide being synthesized was probably more accurate when the aminoacylation and Met were added part way through the incubation time. Therefore, we continued to use this new, interrupted translation protocol.

Still hoping to improve the signal to noise ratio, we wondered if the aa-tRNA pellet was drying completely in 5 minutes. Even a small amount of ethanol left over from precipitating the aa-tRNA could inhibit translation. In an effort to dry the pellet completely, we decided to dry it for sixty minutes (Figure 3.12). Additionally, since it would greatly facilitate our experiments to be able to do aminoacylations a day or two before a translation, we decided to assess if aa-tRNAs were stable at -80 ˚C. We tested a variety of variables including length of storage, and storing them as a pellet or resolubilized in water (Figure 3.13).

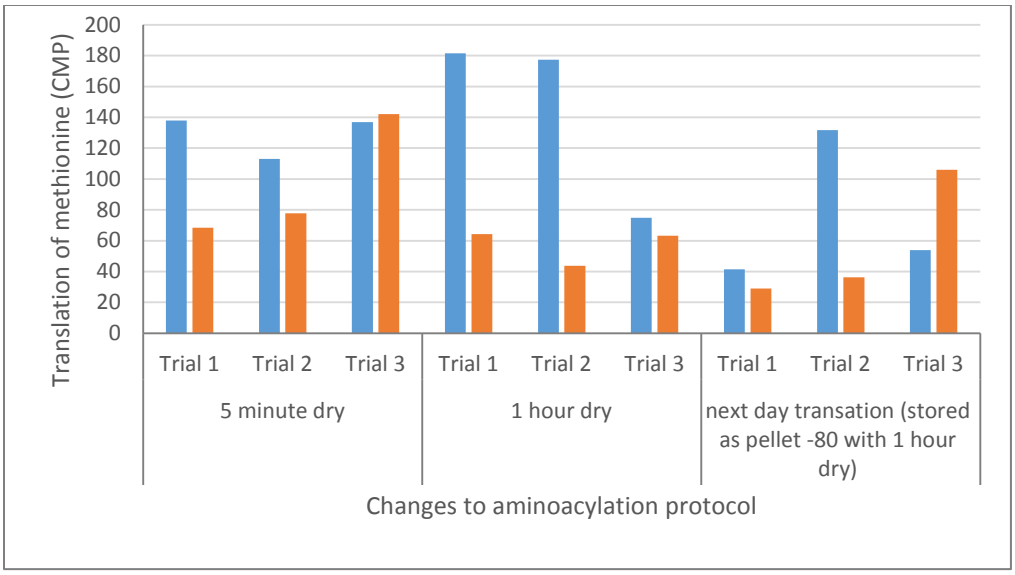

Figure 3.12 Drying aa-tRNA pellet before translation. Translation with the correct MetRS (blue) and with the non-canonical aaRS (orange).

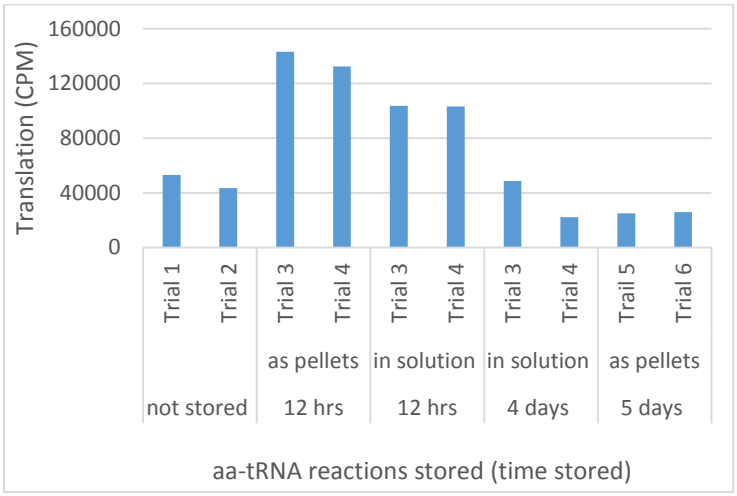

Figure 3.13 Val translation. Assessing storage of aminoacylation reaction after 12 hours, 4 days, or 5 days as a pellet or solution.

The data suggested a sixty minute dry time improved translation. Additionally, it suggested that freezing the pellet also improved translation. As such, we continued to perform aminoacylations and translations on separate days, storing the aminoacylated pellet at -80 ˚C degrees overnight. The data showed we could not store the pellet long-term

as a pellet or in solution without sacrificing activity. Granted, these data were still inconsistent.

We tried to optimize the concentration of aaRS; however, these data had very high background again (Figure 3.14). Consequently, we repeated the experiment and encountered a different problem: scintillation counts were too low; no translation was occurring (Figure 3.15).

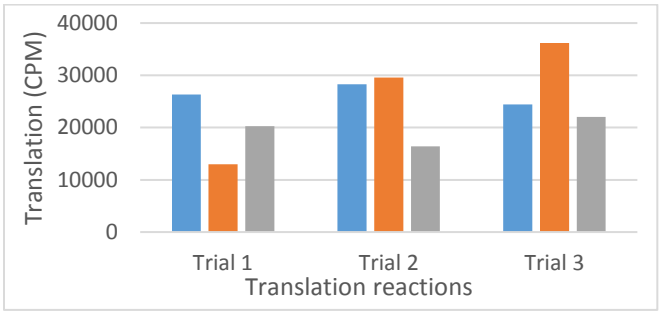

Figure 3.14 Val translation failed with diluted RS. Translation with the correct ValRS diluted 1:4 (blue) and with the non-canonical MetRS diluted 1:4 (orange) and no aaRS (only buffer) (gray).

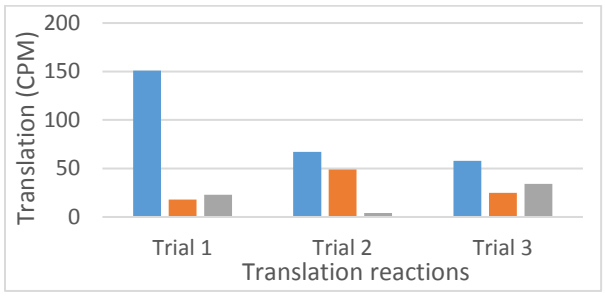

Figure 3.15 Repeat 11.04.12 experiment with Met translation. Translation with the correct MetRS diluted 1:4 (blue) and with the non-canonical ValRS diluted 1:4 (orange) and no aaRS (only buffer) (gray).

3.2.3.3 Experiments stop working

Since we had been constantly changing the protocol, we returned to previously used procedures which we knew worked, but translation was not occurring regardless of which procedure we used (Figure 3.16).

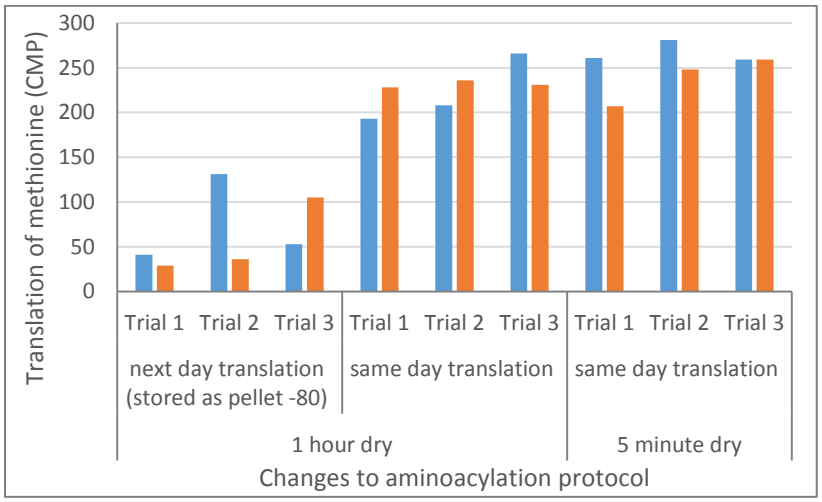

Figure 3.16 Met translation failed. Tried varying protocol. Translation with correct MetRS (blue) and with non-canonical ValRS (orange).

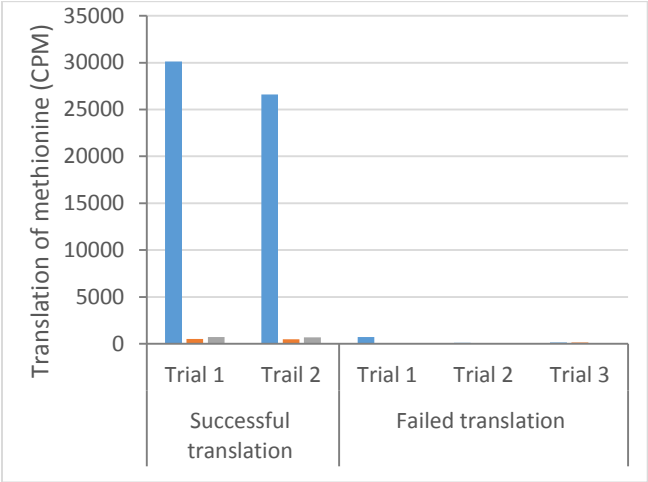

Figure 3.17 Comparing successful translation on left and failed translation results on right. Translation with all necessary components (blue bars) and translation lacking a cognate synthetase (orange bars) is shown.

Comparison of these scintillation values representing successful translation (from 10.19.12) to scintillation values representing unsuccessful translation (from 11.09.12)

confirmed no translation was occurring (Figure 3.17). Radiation counts indicating successful translation should be >10,000 CPMs, but these values were under 200 CPMs.

The simplest explanation was that the PURE translation kit had gone bad due to age, multiple freeze-thaw cycles, or contamination. In response we compared our translation kit (less tRNA) with another translation kit (less EF-Tu), which was successfully generating data, but the results were unchanged (Figure 3.18).

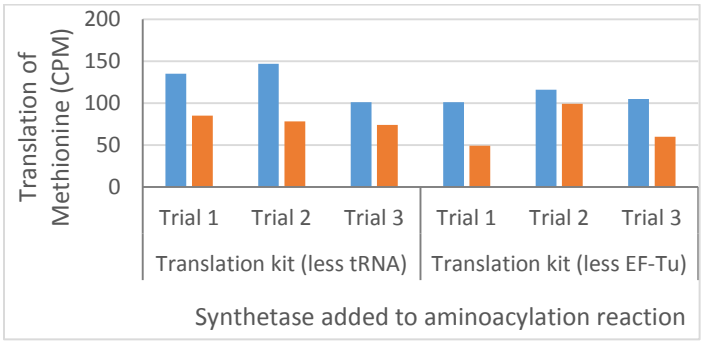

Figure 3.18 Failed Met translation. Comparing translation kits. Aminoacylation with cognate RS (blue) and non-cognate RS (orange).

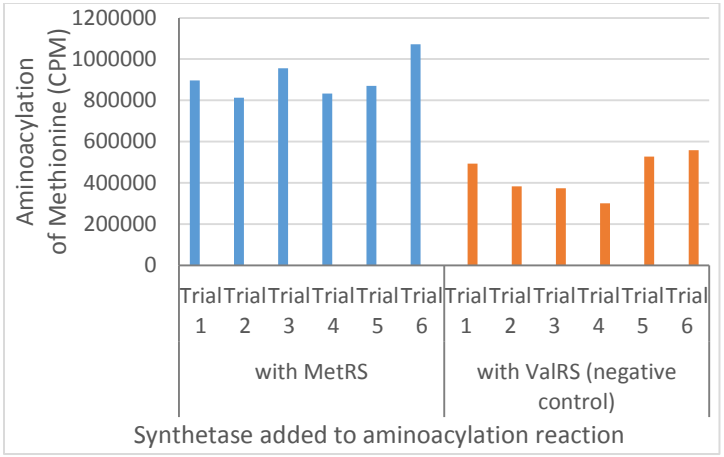

Figure 3.19 Met aminoacylation. Comparing ratio of successful aminoacylation (blue bars) with unsuccessful aminoacylation (orange).

Since the problem was not an expired translation kit, we tested the aminoacylation reaction. Aminoacylation was also not occurring. The ratio of aminoacylation with the

cognate MetRS compared to aminoacylation with the non-cognate ValRS was low (Figure 3.19). This ratio should be at least 10 (Figure 3.20).

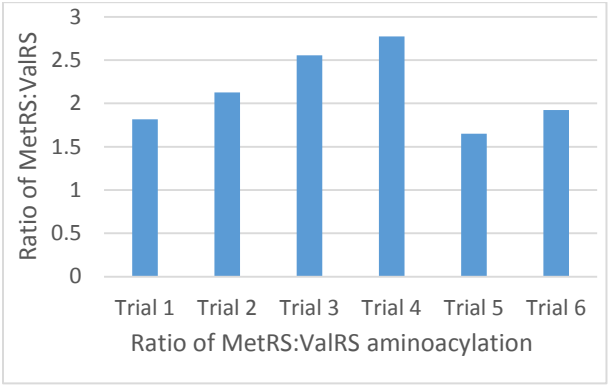

Figure 3.20 Comparing ratio of successful aminoacylation (blue bars) with unsuccessful aminoacylation (orange). Typically ratio should be 6 to 8.

Since we knew at least part of the problem was with the aminoacylation, the most likely culprit was a stock or reagent that had gone bad. We made new, fresh stocks of ethanol, HEPES, dATP (which we moved to the -80 ˚C freezer instead of the -20 ˚C), $MgCl_2$, KCl, PPase, DTT (from the Payne Lab), and new deacylated tRNA.

In duplicate, we performed the experiment carefully following the exact procedure used the last day the experiment worked. The data seemed to suggest the old stock solutions were working, but the new stock solutions were not. We repeated the experiment only through aminoacylation, but the results were inconclusive. Since we were still unsure if the stocks were the issue, we repeated the experiment comparing old and new stocks with each of three scientists working only on her own experiments, but no experiments were successful (Figure 3.21). All ratios were below 10 (Figure 3.22).

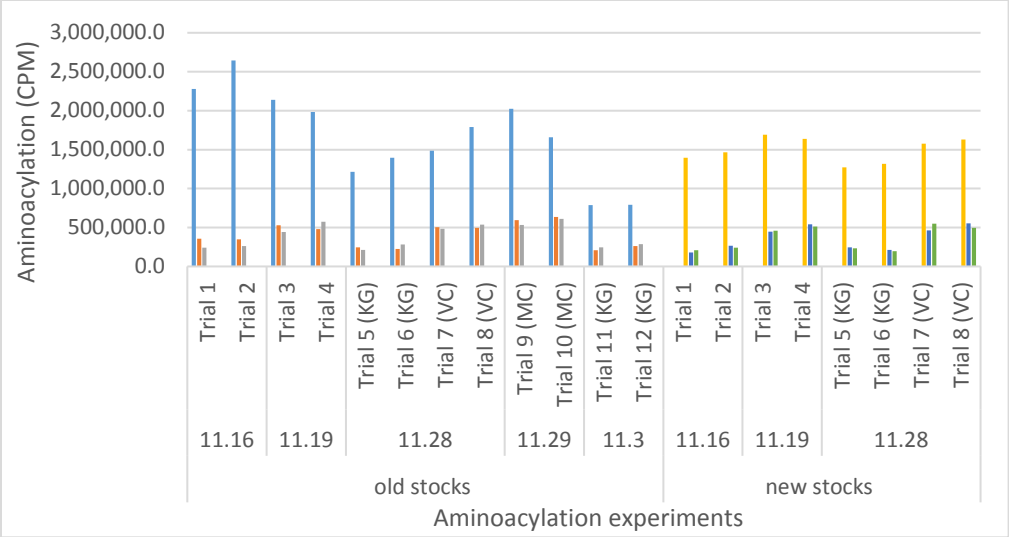

Figure 3.21 Met aminoacylation with old and new reagent stocks. Date of experiment listed with experimenter's initials. Cognate RS (blue/yellow), non-cognate RS (orange/navy), no RS (gray/green).

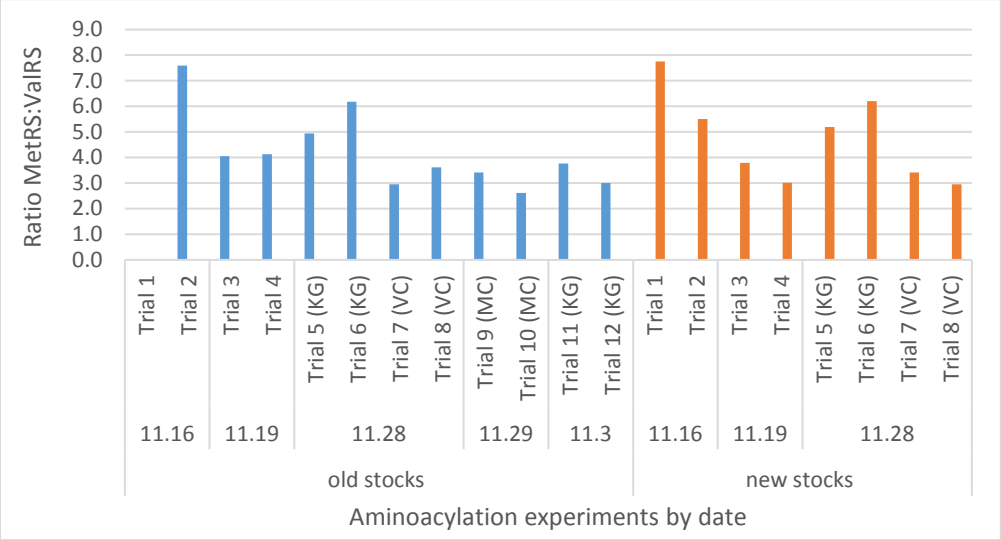

Figure 3.22 Ratio of aminoacylation with cognate (MetRS) compared to negative control (with ValRS). Experiments performed with old (blue) and new (orange) reagent stocks. Date of experiment listed with experimenter's initials.

3.2.3.4 Aminoacylation starts working again

We finally decided to perform an experiment selecting old or new batches of reagents and stocks based on observations and background knowledge. We chose the old

HEPES buffer because we knew it had worked before. We used fresh stocks of $MgCl_2$ and KCl because we started using molecular grade water part way through the fall and we knew the newer stocks had been made with molecular grade water. We chose to use ATP from the -80 ˚C because we thought the ATP stock in the -20 ˚C might have precipitated before it had been aliquoted. We used a fresh batch of deacylated tRNA and new ppase because the ppase was usually liquid in the -20 ˚C and the old ppase had frozen solid. We used the original protocol from the last day the experiment had worked. We resolubilized the aminoacylated reactions in water or in 1mM sodium acetate (pH ~4). We performed each experiment in triplicate. The experiment was successful. However, we could not definitively explain why it had started working. We decided to try to recreate the bad results we were getting when the aminoacylation was not working by using different stocks of HEPES buffer, dATP, and PPase. Unexpectedly, these experiments were also successful (Figure 3.23). All results indicated the aminoacylation had been successful with an average signal to noise ratio of 20:1, possibly the best ratio we had achieved (Figure 3.24).

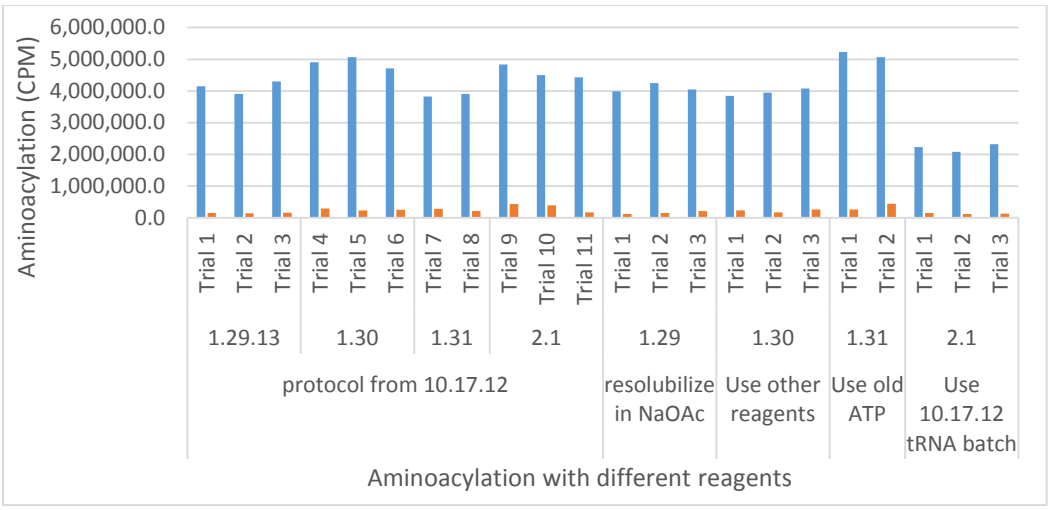

Figure 3.23 Successful Met aminoacylation. Acylation with cognate MetRS (blue) and non-cognate ValRS (orange). Experiments listed with date.

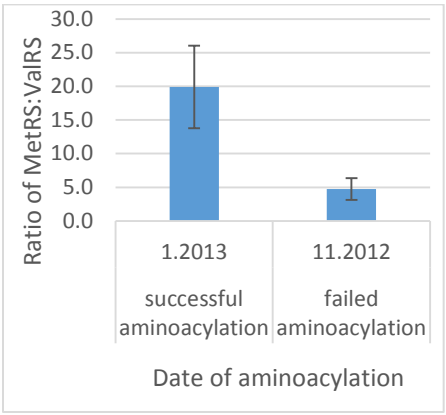

Figure 3.24 Ratio of aminoacylation with cognate (MetRS) to non-cognate (ValRS).

We were ultimately unable to intentionally sabotage the experiment; however, we theorized the culprit may have been a bad batch of tRNA. Since we had used up that batch, we were unable to confirm this theory, but during a discussion with Dr. Loren Williams, he mentioned his lab often had problems with tRNA and they always ran new purchases out on a gel to make sure they had intact tRNA and not only fragments. Supporting this conclusion, aminoacylation rates varied with different batches of tRNAs.

In light of these experiments, it became clear the *in vitro* assay had challenges that were not going to be easily addressed. As such, we decided to initiate an *in vivo* assay as well.

CHAPTER 4. OPTIMIZING THE *IN VITRO* ASSAY FOR EF-TU

Herein we discuss optimization of the translation protocol specifically for work with EF-Tu variants. While similar, this data does not repeat data presented in other chapters. The *in vitro* translation experiments were performed with a kit that lacked EF-Tu and amino acids which is similar to but different from the kit used in the previous chapter. We cannot necessarily assume the *in vitro* translation kit without EF-Tu will react the same as the kit without tRNAs and any differences are critical for us to note.

4.1 Optimizing assay

4.1.1 Experimental protocols

The *in vitro* assay used a custom PURE translation kit (NEB) from which EF-Tus and amino acids had been withheld. Translated peptides had an S-35 radioactive tag on the initiating fMet and were purified via affinity chromatography on Ni-NTA columns. As a result, the quantity of translated peptide is directly related to scintillation (CPM). Although this assay ultimately posed too many challenges to be overcome, it had a number of advantages over *in vivo* assays. First, host organism viability was not a concern. Second, we could characterize variants' activity with practically any ncAA with no regard for its permeability across the cell membrane and with minimal concern for bioavailability, toxicity, or solubility. Third, we could control (at least in theory) the presence of endogenous translation components.

4.1.2 Key achievements and challenges

The *in vitro* translation assay posed a number of challenges, the most pressing of which was high background. We worked to optimize the DNA template (including codon selection and concentration), amino acid concentration, the ratio of cold to radiolabeled methionine, and reaction volume. By optimizing the protocol, we successfully reduced background by reducing contaminating amino acids and tRNAs. We also improved reproducibility by improving consistency of ribosome allocation in the assay. Via mass spectrometry we also confirmed that codon selection is important. We attempted to pursue a theory of codon:anticodon mispairing in *in vitro* translation but results were inconsistent, and this was abandoned in favor of prioritizing the main focus of the project, EF-Tu engineering.

Challenges that were not readily overcome included a highly thermo-unstable radio-compound. This was eventually exchanged for a more thermostable formulation. We also witnessed extremely low translation rates for ancestral proteins and low reproducibility which made it difficult to draw conclusions (this data was generated by Dr. Cole). An additional challenge was normalizing EF-Tu concentrations. Since our purified EF-Tus contained contaminating peptide product, Bradford protein assays were not applicable and it was difficult to quantify the concentration of only the EF-Tu variant.

4.1.3 Failed control reactions

Our first challenge was getting consistently reliable results from control experiments, especially negative control reactions. Alongside each experiment, we ran a series of negative controls. The negative controls involved one of several options: withholding one amino acid required for peptide synthesis, withholding all amino acids from the translation reaction, or withholding the DNA template from the reaction. We

expected all these negative controls to have similar, low signal, but this was not the case. When we withheld all amino acids or the DNA template from translation, the results matched our expectations, and we saw very low levels of translation However, when we withheld only one necessary amino acid (in this case valine) we saw high translation levels of peptide that were very similar to the positive control (Figure 4.1). These results suggested that peptide translation was happening even when we withheld a necessary amino acid from the translation system. Initially, we used the peptide $MVVMH_6$ in which valine was the amino acid of interest.

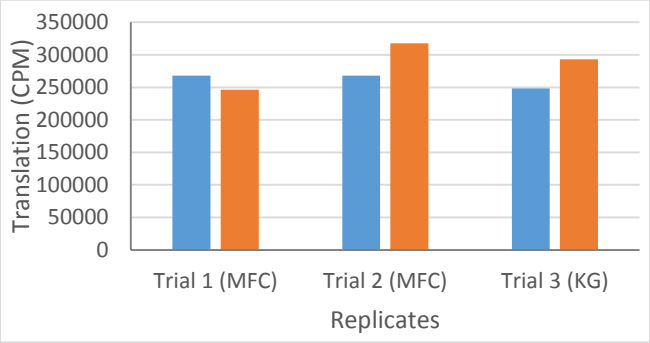

Figure 4.1 Transformation of $MVVMH_6$ plasmid. Background (orange bars) in which the amino acid valine has been withheld is equal to positive control (blue bars) in which all necessary amino acids have been added.

We decided to tackle the problem using multiple approaches. Our first strategy was to optimize the amino acid concentration. We thought one possible explanation was that when valine was withheld, another amino acid was being translated at that codon in lieu of valine (Figure 4.2). We found that diluting the amino acid concentration did not result in a better signal to background ratio; however, we did learn we could reduce the amino acid concentration by a factor of ten and still have similar results. In these results, our negative control was significantly lower than in some other experiments. However, it returned to being higher like before. As such the issue became both higher background (we hoped for

a ten-fold difference in CMP between signal and background) as well as consistence. As such, this did not completely solve our problem with high background, but did provide additional information.

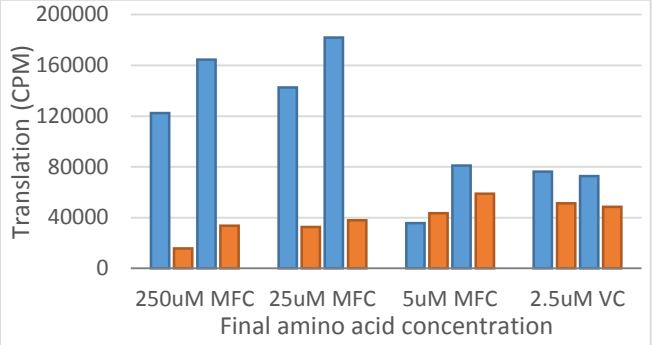

Figure 4.2 Optimizing amino acid concentration. Translation with all amino acids (blue bars) and translation lacking valine (orange bars).

Concurrently, we experimented with using non-radiolabeled methionine in addition to only radiolabeled methionine (Figure 4.3). We found using a mixture of both radiolabeled and non-radiolabeled methionine resulted in a much better ratio of signal to background (Figure 4.4Figure 4.3).

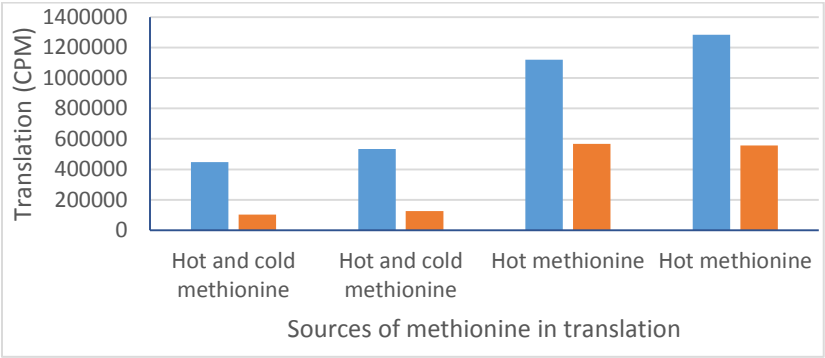

Figure 4.3 Translation either with both hot and cold methionine or with only hot methionine. Blue bars contain all amino acids; orange bars lack valine.

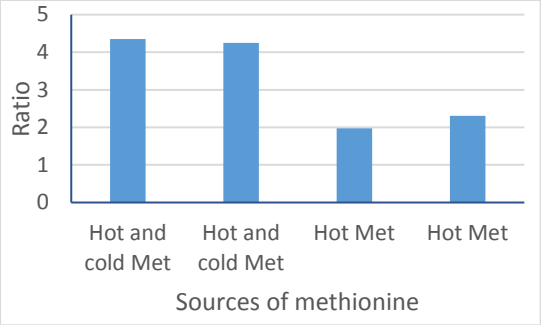

Figure 4.4 Ratio of signal to background with either both hot and cold methionine or only hot methionine.

Another theory was that radioactive methionine was being translated at the valine codon. Were that the case, each peptide in the negative control experiments would have multiple molecules of hot methionine which would increase background. If this were the case and radioactive methionine was going in for the two valines, only one third the number of peptides would need to be translated to give the same signal as our experimental reactions. Additionally, another possibility was that when valine was withheld, the system was not initiating translation at the first methionine, but rather at the second, translating MH_6 instead of $MVVMH_6$. We addressed both possibilities by designing a new peptide, an 11-mer $MVEVRH_6$, but background was still high.

4.1.4 Optimizing DNA template and follow up experiments

We wondered if a full-length protein would also have high background. A full-length protein would have many valines that needed to be translated and would also have different valine codons for the translation machinery to read-through. This experiment seemed to give us accurate results: significant translation occurred with all twenty amino acids and withholding valine resulted in very low yields (Figure 4.5). At this point, we knew our controls worked if we translated a large protein, but our controls failed if we

translated a short peptide. We thought perhaps the problem was the valine codon we chose. Perhaps a different amino acid was going in for valine. We noticed the Val[GTG] and Met[ATG] codons had only one different nucleotide so we theorized that perhaps methionine was going in at the valine codon(s). A mispairing between a $tRNA^{Met}$ anticodon and the valine codon does not follow normal theories of anti-codon:codon wobble pairing; however, because the *in vitro* translation kit does not contain all translation components that are in the cell, it seems possible it may not follow the same rules regarding codon:anti-codon wobble.

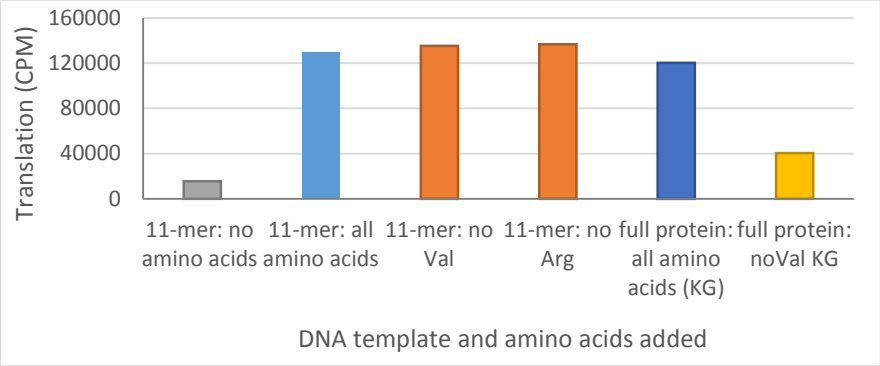

Figure 4.5 Translation of 11-mer $MVEVRH_6$ peptide and EF-Tu protein. Background (orange bars) is equal to signal (blue bar) for peptide. Background (yellow) is as expected relative to positive control (navy bar) for EF-Tu protein.

We tested this theory by generating four new DNA templates, four variants of MVH_6, each with one of valine's four codons (GTT, GTC, GTA, GTG), as well as a control, MAH_6, in which alanine's codon GCA was used (Figure 4.6). Results showed we only got high background with Val[GTG]; when valine was withheld, all other codons showed low translation on par with MAH_6 (for which alanine was withheld). The negative control with no DNA template had the least translation. This supported our theory that

methionine was being incorporated at the valine position when the Val[GTG] codon was used, since the codons differed by only one nucleotide.

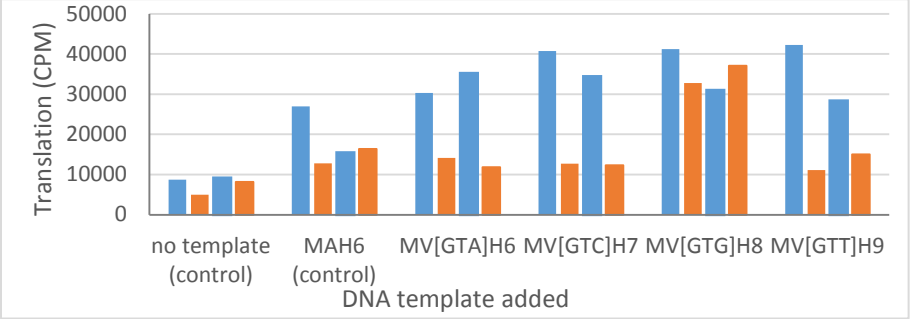

Figure 4.6 Testing different valine codons in DNA template for peptide translation. Template with alanine in lieu of valine is used as a control. Amino acids added were either MVH (blue bars) or MH (orange bars).

Using mass spectrometry, we confirmed that methionine was going in for valine at the Val[GTG] codon. Because *in vitro* translation does not contain all components of *in vivo* translation, it is possible codon:anti-codon wobble pairing is different in vitro than *in vivo*. It is relevant to note that although researchers are familiar with differences between *in vitro* and *in vivo* translation, why we see these differences is not understood; however, observing differences is not unusual. When we translated MV[GTG]H_6 without valine, the resulting peptide had a mass consistent with formyl-MMH_6 and lacked a peak for formyl-MVH_6. However, when valine was included in the reaction, a peak was observed consistent with the mass of formyl-MVH_6 and no peak was observed for formyl-MMH_6. A control reaction using no amino acids showed no peaks consistent with either peptide (Figure 4.7). It is important to note that the *in vitro* translation performed with the radiolabeled Met and the mass spectrometry data are not actually replicates; although we thought so at the time. The radiolabeled Met also contained Val. Because we could not use radiolabeled Met to

prepare peptides for mass spectrometry, these peptides were generated with regular Met and thus did not contain Val.

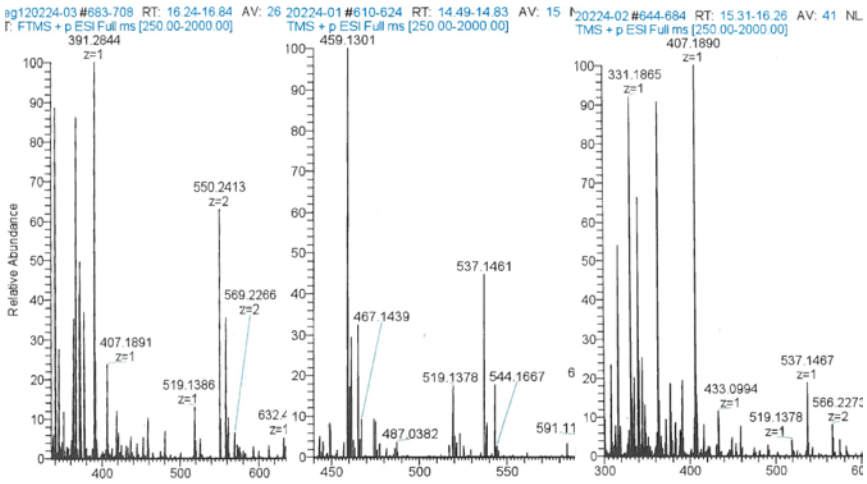

Figure 4.7 Mass spectrometry. Panel A shows a peak at 550 corresponding to fMVH$_6$ (m=1100), panel B is a negative control, and panel C shows a peak at 566 corresponding to fMMH$_6$ (m=1132).

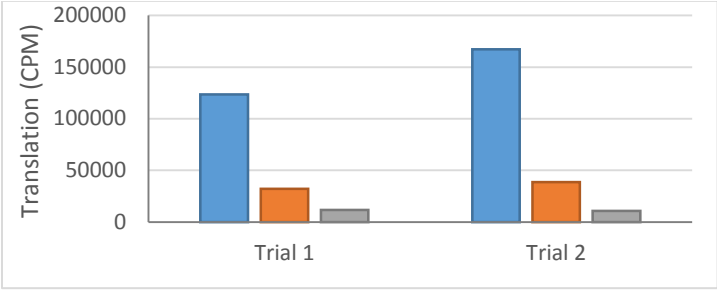

Figure 4.8 Translation of MVEVRH$_6$-2 DNA template in which no codons share more than one nucleotide. Background (gray bars) had no amino acids added, the negative controls (orange bars) lack valine, signal (blue bars) show peptide translation.

Based on the results indicating that the valine codon choice was critical to maintaining low background, we designed a new template, MVEVRH$_6$-2, with no codons that shared more than one base in a particular position: M[ATG]; V[GTT]; E[GAA]; R[CGA]; H[CAC, CAT]). We tested the MVEVRH$_6$-2 DNA template *in vitro* and

observed good reproducibility and low background. Withholding valine resulted in only slightly higher readings than when no amino acids were included at all (Figure 4.8).

4.1.5 Optimizing assay protocol

Now that all experimental controls were working, we worked to improve reproducibility. We tested different DNA template concentrations and learned quantities as low as 17.5 ng per reaction saturated translation (Figure 4.9).

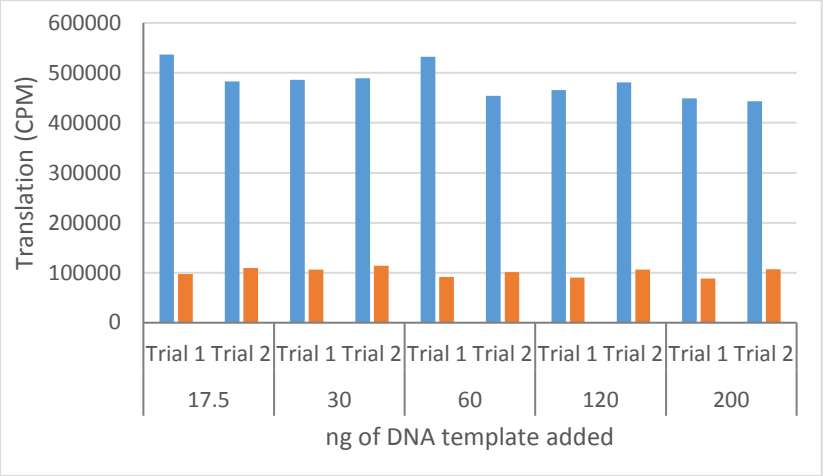

Figure 4.9 Translation with different concentrations of DNA template. Blue bars contain all necessary amino acids; orange bars lack valine.

We noticed that radiation values steadily declined from the first sample through the last. Each day, we ran several replicates simultaneously in separate Eppendorf tubes. Throughout the experiment, the tubes were treated (a solution was added or procedure was performed) in the same order. Even though all tubes were replicates of the same experiment, the first tube consistently had higher radiation levels with levels steadily decreasing through the last tube.

One possible explanation for the observed inconsistency was that some of the samples sat longer at certain points in the protocol while the technician was working with

other samples. To test if the timing was the critical factor leading to inconsistent results, we intentionally allowed the experiment to sit for two minutes at various points in the procedure; however, intentionally pausing the experiment made no difference in the results (Figure 4.10). We would have expected any difference in CPM among samples to more dramatic than usual since we intentionally allowed the samples to sit for longer than they would in a normal procedure.

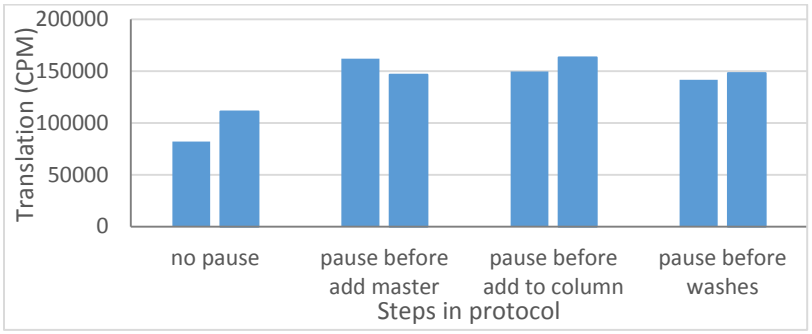

Figure 4.10 Translation paused for two minutes at different points in protocol.

Another possibility had to do with aliquoting a master mix to the tubes. During the protocol, a master mix of the translation solution was made for the day's experiments. This solution was then aliquoted to individual tubes for the various experiments and controls. It was possible the tubes were not getting equal amounts of translation components. Specifically, ribosomes are known to sink in the solution. Since the ribosomes in the *in vitro* translation kit have limited turnover, the number of ribosomes aliquoted to each tube could affect translation rates. More ribosomes would result in more peptide production reflected by higher radiation values. To test this theory, we implemented adjustments in the protocol to ensure a homogeneous suspension of the translation components when we aliquoted the master mix.

To further improve reproducibility, we tested increasing the reaction volume (from 5uL to 6uL total volume) (Figure 4.11). Pipetting larger volumes could help improve pipetting accuracy. Increasing the volume did not show any effect on the results and allowed us to add larger volumes of some solutions.

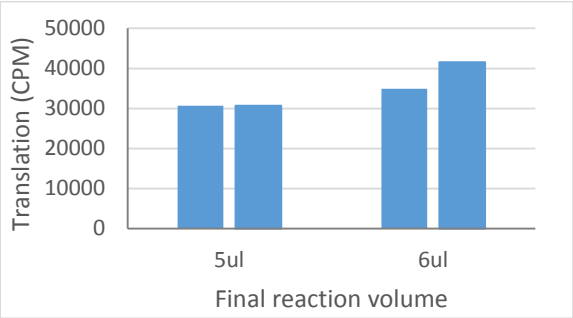

Figure 4.11 Translation with different volumes.

Although our translation kits were ordered without amino acids or EF-Tu, it was still possible some contaminating aa-tRNAs or amino acids were in the NEB kit solutions. We thought these contaminating amino acids might result in an increase in background signal. Adding no amino acids, we ran an experiment where we stopped translation a different time points. We found that despite withholding amino acids from translation, samples still showed translation during the first sixty minutes of the reaction (a normal reaction time was three hours) (Figure 4.12). This indicated it took approximately sixty minutes for any contaminating amino acids to be translated into peptides. This led to an adjustment in the experimental protocol; we decided to use an interrupted incubation period. We would incubate the solution for sixty minutes, then add radioactive methionine and finish the incubation. We found this strategy significantly reduced background. However, because it also reduced signal, the signal to background ratio was unaffected.

We decided to continue using this protocol because we thought even though the signal was lower, the results probably more accurately represented amount of peptide synthesized.

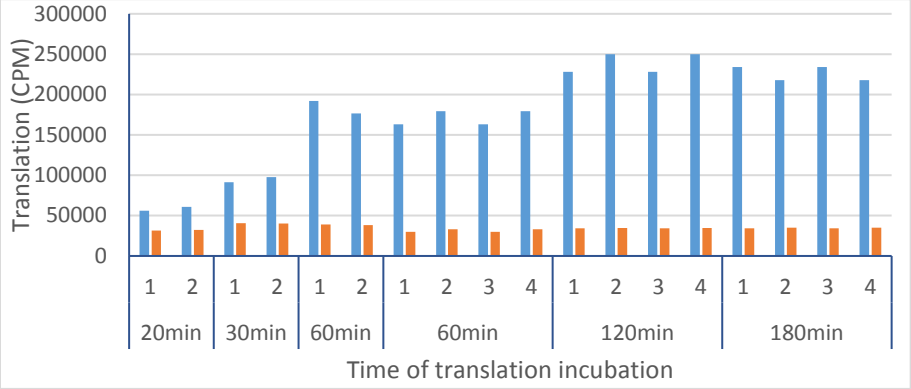

Figure 4.12 Timecourse of translation. Measure of peptide production relative to time of translation. Blue bars contain all needed amino acids; orange bars lack valine.

4.1.6 Different codons in the DNA template

After we optimized the translation protocol, we were intrigued by the idea that codon:anticodon wobble pairing followed different rules *in vitro* than *in vivo*. Recall that we found Met[ATG] was able to be translated at a Val[GTG] codon. The Met-tRNA^{Met} was mispairing in the first position, not the last as with traditional wobble pairing. Curious to know if a general rule could be identified, we synthesized additional DNA templates exploring codons that had two nucleotides in common with Met[ATG]. Codons explored included Ile[ATT, ATC, ATA], Leu[CTG, TTG, TTA, CTT, CTC, CTA], Lys[AAG, AAA], Thr[ACG, ACT, ACA, ACC], and Arg[AGG, CGT, CGC, CGA, CGG, AGA]. Each codon was inserted into a MXH$_6$ DNA template. Since these DNA templates were very short, primers were ordered, amplified, digested, and ligated into a digested expression vector.

We first tested leucine codons, CTG, TTG, TTA, and CTA (Figure 4.13). We expected methionine would go in for CUG and we would see translation even if leucine were withheld from the reaction. However, when we translated the peptide MLH_6, all leucine codons had low background. Although Leu[CUG] had higher background, it was not high enough to indicate significant Met incorporation.

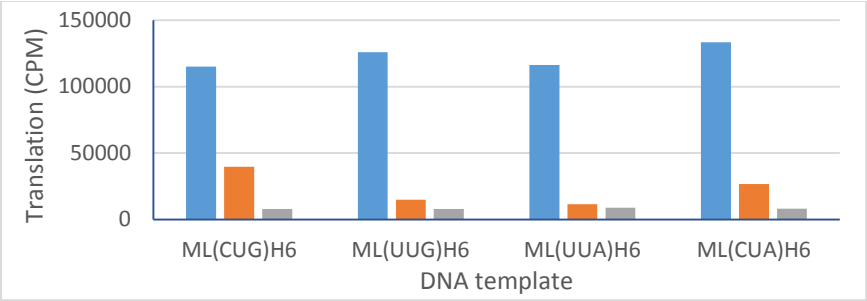

Figure 4.13 Translation of DNA templates containing different Leu codons. Blue bars are samples that contain Leu, orange bars lack Leu, and gray bars have no amino acids added.

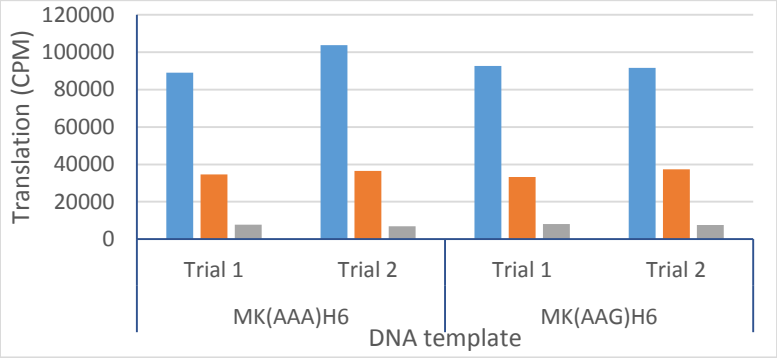

Figure 4.14 Translation of DNA templates containing different Lys codons, AAA and AAG. Blue bars contain Lys, orange bars lack Lys, and gray bars have no amino acids added.

We also tested lysine which we expected to have methionine misincorporation at the AAG codon; however, the AAG and AAA codons had very similar results. We found

no evidence of methionine misincorporation when lysine was withheld from the reaction (Figure 4.14).

Later, we learned that the radiolabeled methionine contains valine and alanine in an approximate 2:1 molar ratio (valine or alanine to methionine). This explains, in part, why we struggled with the *in vitro* assay. The decision to use valine analogs was perhaps uninformed despite the advantage that $tRNA^{Val}$ does not require posttranslational modifications (the driving force behind the decision to test valine analogs).

4.2 Follow up with EF-4A

Following the experiments listed above, the *in vitro* assay was dropped due to a litany of challenges including consistently poor reproducibility and the inability to normalize EF-Tu concentrations (see below for full explanation). However, after generating data indicating EF-4A was a viable variant, we returned to the *in vitro* assay to assess EF-4A's capabilities with other ncAAs. It is relevant to note, the radio-isotope used was highly thermo-unstable and as such, over the course of three months, we were unable to generate data. In detail, the radio-compound was highly volatile which resulted in significant contamination issues. Each decontamination effort required a freeze-thaw of the compound which resulted in complete degradation each time.

4.2.1 New DNA templates

As stated, after finding out variant EF4A had activity with Sep, we moved to the *in vitro* assay to assay its activity with other ncAAs that are, to varying degrees, compatible with wild type tRNA and aaRSs.[32, 51] As such, we hoped that EF4A would be able to

improve upon ncAA translation. We were looking for increased translation when EF4A was added to translation, as compared to ncAA translation with only wild type EF.

Table 4.1 New DNA templates synthesized.

Amion acid targeted	codon used	DNA template
Ala	GCA	MAEARH6
Arg	CGA	MRVREH6
Glu	GAA	MEVERH6
Ile	ATA	MIEIRH6
Ile	ATC	MIEIRH6
Leu	CTA	MLELVH6
Lys	AAA	MKVKRH6
Phe	TTT	MFEFRH6
The	ACC	MTETRH6
Trp	TGG	MWEWRH6
Tyr	TAC	MYEYRH6
Tyr	TAT	MYEYRH6
Val	GTT	MVEVRH6

In lab we had ncAAs which could go in for either Ala, Arg, Glu, Ile, Leu, Lys, Phe, Thr, Trp, Tyr, or Val.[32, 51] We synthesized appropriate DNA templates using digestions and ligation (Table 4.1). Using only NEB EF-Tu, we first tested valine and alanine analogs, then glutamate, lysine, leucine, and arginine (Figure 4.15, Figure 4.16). We tested two different codons for Ile and Tyr so see if one of the codons had better signal to background ratio (Figure 4.17). Finally, we tested Phe, Thr, and Trp. For the Thr template, results were inconclusive so the experiments was repeated several times (Figure 4.18). Glu template was also replicated (Figure 4.22). We typically hope for values above 100,000 CPM to indicate successful translation. We hope for controls to be ten times lower.

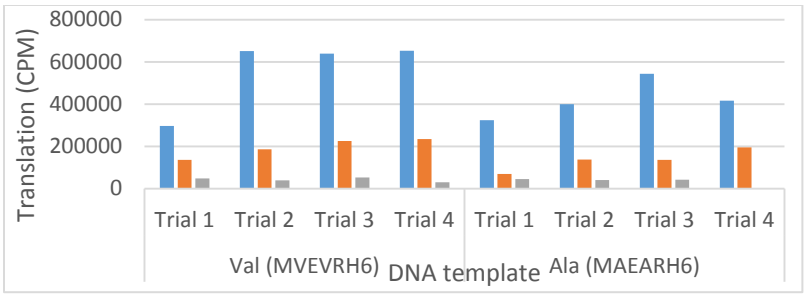

Figure 4.15 Checking control reactions when using new DNA templates $MVEVRH_6$ and $MAEARH_6$. Bars show translation with all necessary amino acids (blue), lacking valine or alanine, respectively (orange), and lacking all amino acids (gray).

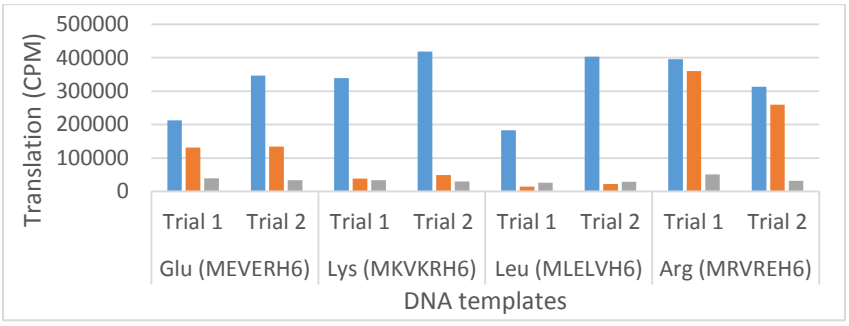

Figure 4.16 Checking control reactions when using new DNA templates. Bars show translation with all necessary amino acids (blue), lacking the target amino acid (orange), and lacking all amino acids (gray).

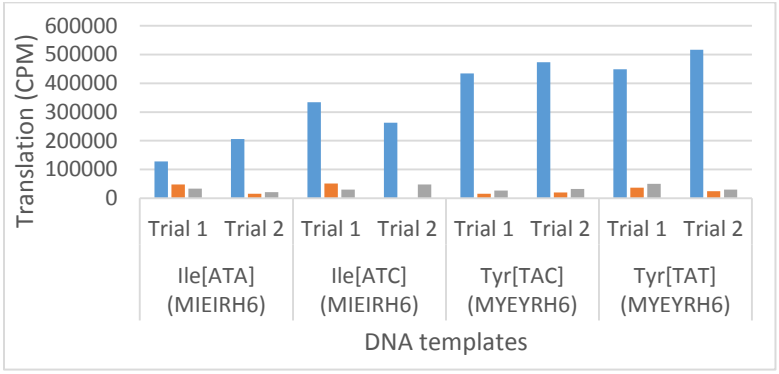

Figure 4.17 Checking control reactions when using new DNA templates. Bars show translation with all necessary amino acids (blue), lacking the target amino acid (orange), and lacking all amino acids (gray).

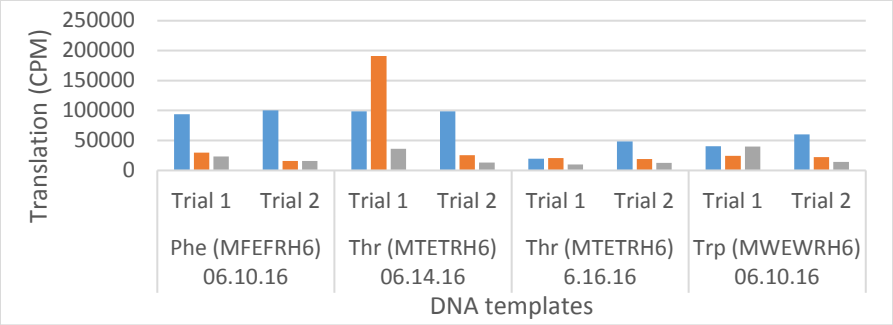

Figure 4.18 Checking control reactions when using new DNA templates. Bars show translation with all necessary amino acids (blue), lacking the target amino acid (orange), and lacking all amino acids (gray).

4.2.2 Continued protocol optimization

Under Dr. Cole's direction, it had been decided that when pipetting such small volumes, the protocol would be to not wet the pipet tip and to expel the solution, pressing through to the second stop; however, we were not certain if this was the best technique. We ran experiments in which we expelled to the first stop with a dry tip, expelled to the second stop with a dry tip as we had been doing, and also first wet the pipet tip and then expelled to the first stop (Figure 4.19). During the time we had a number of issues with the assay not working (likely due to degradation of the radio-compound) but despite messy data, we decided wetting the tip first, then expelling to the first stop gave the best reproducibility. We based this analysis on the first set of data since the second set had really low translation rates, suggesting the radio-compound had degraded and thus we did not completely trust the results (Figure 4.20). Additional changes to the *in vitro* translation protocol included wetting the pipet tip when washing the Ni-NTA column with stop buffer and when eluting with imidazole. We also added a dry spin after the second wash with stop buffer and before the elution step.

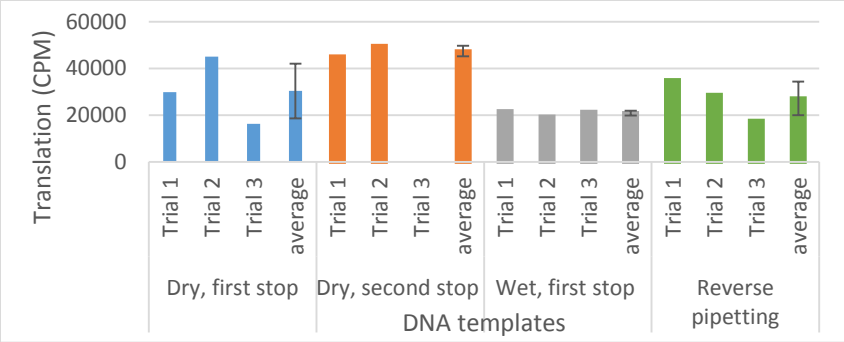

Figure 4.19 Different pipette techniques. Bars show translation with all necessary amino acids. Colors are simply to guide the eye to replicates/averages.

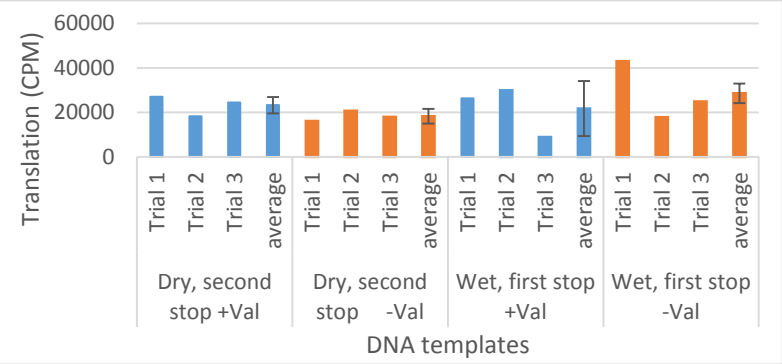

Figure 4.20 Different pipette techniques. Bars show translation with all necessary amino acids (blue) or missing valine (orange).

EF-Tu variants were diluted and stored in a buffer solution. We tested if this buffer inhibited translation (as compared to water). The data translating the Val DNA template was messy so it was impossible to come to a firm conclusion (sometime standard deviations for reactions with water were smaller than reactions with the buffer and sometimes they were larger). We decided to try the experiment again with the Lys DNA template but values indicated no translation occurred (Figure 4.21).

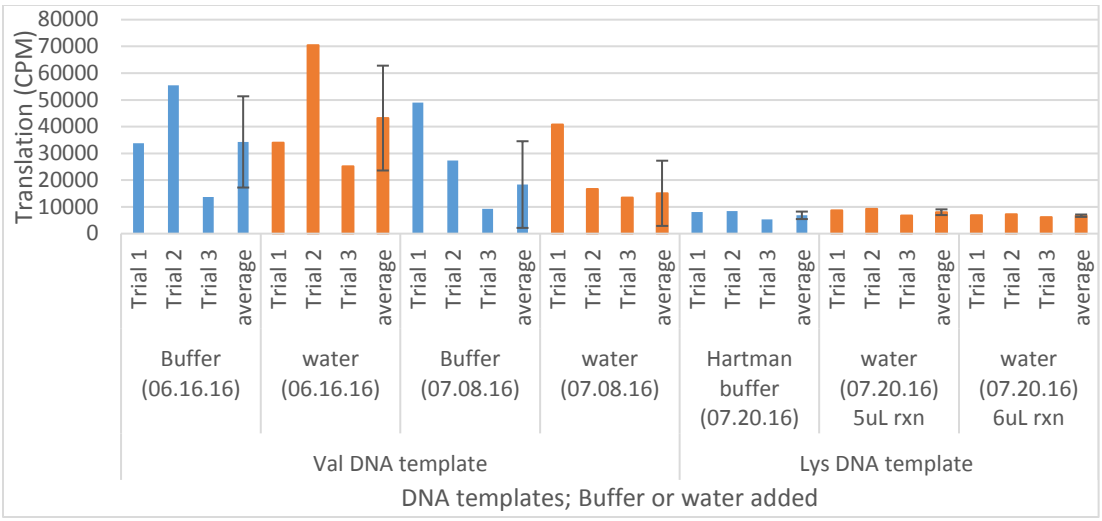

Figure 4.21 Assessing effect of buffer (blue bars) on translation as compared to water (orange bars).

4.3 Radio-compound formulation

In detail, our radioactive Methionine, L-[35S] from Perkin Elmer (product number NEG009T005MC) is highly volatile because it is stored in beta-mercaptoethanol. As such, it must be stored in a container with activated charcoal in the container as well as activated charcoal in dishes in the freezer where it is located. It is also highly temperature sensitive and if warmed above -20 ˚C will degrade to methionine sulfoxide-[35S]. Storage at -80 ˚C is preferable, but storage at -20 ˚C is acceptable (which is what was used).

However, Perkin Elmer developed another product, EasyTagTM Methionine, L-[35S] (produce number NEG709A500UC) which is stored in a stabilized aqueous solution. This formulation has added stabilizers as well as a blue dye, tricine, and β-mercaptoethanol. It is stable at room temperature and can be stored at 4 ºC. For long-term storage, -20 ºC is suggested to reduce risk of microbial contamination.

These products have identical concentration of methionine. The concentration of radiation is very similar (11 mCi/mL and 10.25 mCi/mL for NEG009T005MC and NEG709A500UC, respectively).

It seemed translation with the new methionine worked well, however, it did not match results from earlier experiments (Figure 4.22).

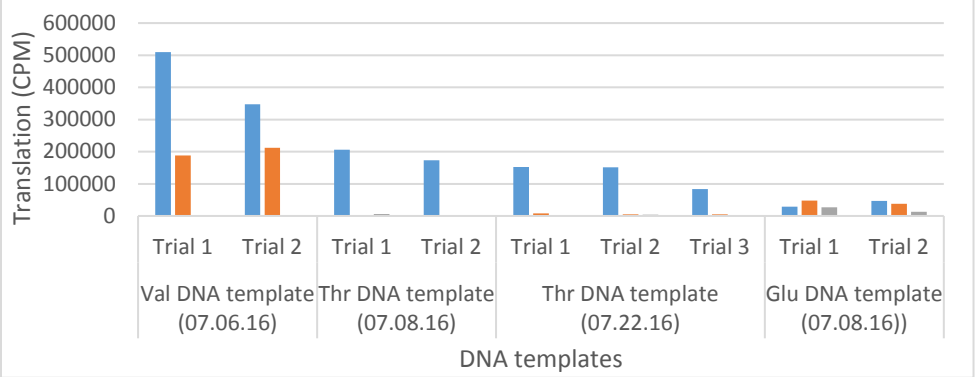

Figure 4.22 Translation with new, thermo-stable methionine. Bars show translation with all necessary amino acids (blue), lacking the target amino acid (orange), and lacking all amino acids (gray).

4.4 Normalizing EF-Tu concentration

In general it is relevant to note that it was difficult to normalize concentrations of EF-Tus purified in house. In order to use our purified EF-Tus, we had to address the problem of not being able to quantify the concentration of our EF-Tu variants. One problem we ran into was standardizing the concentration of our EF-Tu variants. We couldn't use a Bradford assay to determine the EF-Tu concentration because our EF-Tu variants had protein contamination from an unknown source and Bradford assays measure total protein concentration. Since we had data that indicated a wide range of EF-Tu concentration was acceptable in the *in vitro* assay with no effect on results, we decided to not expend additional effort on normalizing the EF-Tu variants. However, depending on the direction

the project take in the future, we could perform a protein gel purification step followed by a Bradford assay. In this strategy, we would run the EF-Tu variant on a protein gel, excise the EF-Tu band, dissolve the gel, and resolubilize the protein. Then, it may be possible to perform a Bradford assay and get an accurate EF-Tu concentration of our variants.

Alternatively, using a hexahistidine tag often purified contaminating proteins. We could purify the EF-Tu variants again and wash the affinity column using a buffer with a higher concentration of imidazole (e.g. 200mM). Some product would be lost, but the sample would likely be significantly more pure.

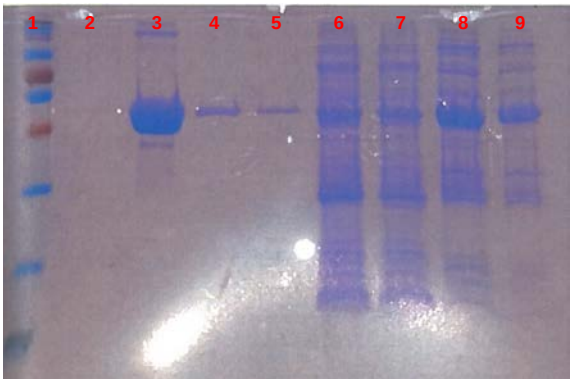

Figure 4.23 Protein gel of EF-Tu variants at difference concentrations. Columns show (1) ladder (unfortunately, company is unknown); (2) the Hartman buffer (empty of proteins); NEB EF-Tu solution (3) at full centration, (4) 1:20x dilution, and (5) 1:40x dilution; in-house purified EF-4A (6) at full concentration and (7) 1:2 dilution; and in-house purified wild type EF-Tu (8) at full concentration and (9) 1:2 dilution.

However, since translation remained consistent over a wide range of EF-Tu concentration, we decided to move on with translation experiments and quantify EF-Tu concentration later if it seemed more important. We knew that the NEB EF-Tu can be diluted 1:20 with no change in translation. As such, we decided to try to visually estimate the EF-Tu concentration on a protein gel. Run on a protein gel, the NEB EF-Tu diluted

1:20 looked about the same as the EF4A diluted 1:2 and wild type EFcoli looked somewhat more concentrated than both (Figure 4.23).

CHAPTER 5. ASSAYING EF-TU AND TRNA VARIANTS *IN VIVO*

5.1 Summary

The first part of this chapter is excerpted from a manuscript. The latter part of the chapter details procedure, components, protocols, troubleshooting, and additional data from other libraries which did not lead to publications.

5.2 Successful ncAA incorporation

5.2.1 ncAA-compatible EF-Tu variants

To characterize the EF-Tu REAP-derived library, we used an amber codon suppression assay requiring the co-translational insertion of an ncAA at a premature amber stop codon. The target gene, chloramphenicol acyltransferase (CAT), contains an amber mutation at the permissive D112 position. CAT confers antibiotic resistance to *E. coli* resulting in an assay that directly correlates ncAA incorporation with cellular survival reported as half the maximal inhibitory concentration (IC_{50}). Rates of survival above wild type EF-Tu (EF-coli) indicated the REAP-engineered EF-Tu variant could translate the ncAA with greater efficiency than EF-coli.

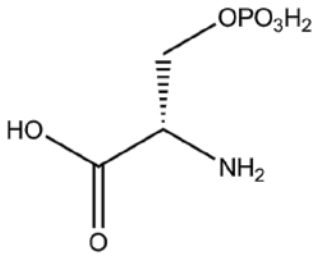

Figure 5.1 Chemical structure of o-phospho-L-serine.

Phosphoserine (Sep) was a strong ncAA candidate for an engineered EF-Tu, because it is incompatible with wild type EF-Tu (Figure 5.1). To overcome this barrier, previous work developed an orthogonal triplet consisting of $tRNA^{Sep}$, SepRS, and EF-Sep to enable co-translational insertion of Sep.[17] This engineered triplet provided a platform for assessing the substrate compatibility of our modified EF-Tus irrespective of whether Sep or another amino acid was incorporated at the amber mutation. Our eight EF-Tu variants were assayed in combination with the Sep-OTS, specifically SepRS and $tRNA^{Sep}$ (Figure 5.2).

A

variant	amino acid position 13	63	64	65	66	67	68	78	214	215	216	227	230	259	261	262	263	272	273	274	275	276
E. coli	N	N	T	S	H	V	E	H	I	E	D	V	R	E	F	R	K	E	N	V	G	V
EF-4A		A									A						A		A			
EF-4B			A					A							A			A				
EF-4C				A						A		A										A
EF-4D	A		A		A									A								
EF-4E						A			A				A								A	
EF-4F							A	A								A			A			
EF-8A	A			A		A				A			A		A	A			A			
EF-12A		A	A		A			A	A		A	A		A			A	A		A		A

B

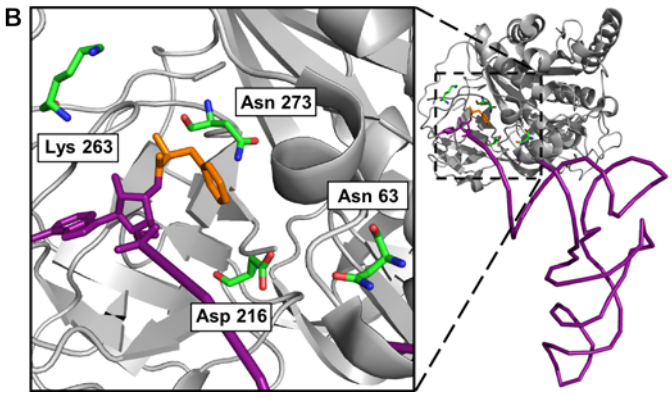

Figure 5.2 Characterization of REAP library variants. (A) Mutations made to generate REAP library. Protein variants are named by the number of mutations (four, eight, or twelve) followed by letters to distinguish among variants with four mutations. Sequence of wild type E. coli EF-Tu is shown for reference. (B) EF-Tu (gray) with amino acid residues mutated in variant EF-4A (inset). Protein is complexed with phenylalanine (orange) and tRNA (purple).

Of the REAP-designed EF-Tu variants, variant EF-4A resulted in the highest levels of CAT translation. Other EF-Tu variants had limited compatibility with Sep, resulting in CAT translation that was either on par with or less efficient than EF-coli (Table 5.1). Although CAT expression was higher with EF-Sep than EF-4A, Sep translation, specifically, was not the aim of this effort. Rather, EF-4A presented a promising lead because it exhibited ncAA substrate compatibility without being specifically engineered for a particular ncAA. This behavior make EF-4A a good candidate for further analysis and development.

Table 5.1 IC_{50} values derived from REAP-derived EF-Tu variants. P-values are relative to EF-coli.

EF-Tu variant	n (replicates)	mean	st dev	p-value
EF-4B	3	0.0088	0.014	0.00001
EF-4C	3	7.74	0.22	0.910813
EF-4D	3	2.13	1.13	0.00001
EF-4E	3	1.29	1.28	0.00001
EF-4F	3	1.89	1.40	0.00001
EF-8A	5	1.11	1.63	0.00001
EF-12A	3	4.71	0.75	0.000669
EF-K263A	3	5.90	0.28	0.015452
EF-Sep	5	26.87	5.83	0.00001

Variant EF-4A had four point mutations: N63A, D216A, K263A, and N273A (Figure 5.2). To determine which of these mutations contributed to ncAA compatibility, four single-mutation EF-Tu variants were generated: EF-N63A, EF-D216A, EF-K263A, and EF-N273A. Of these variants, EF-D216A showed improved efficiency of ncAA translation compared to EF-coli and EF-4A (Figure 5.3).

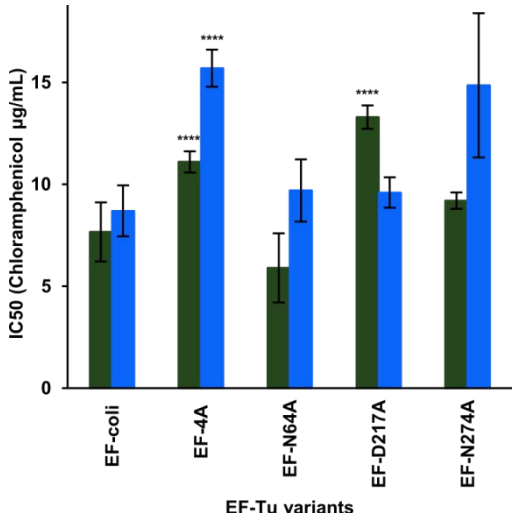

Figure 5.3 Characterization of single-mutant EF-Tu variants. *In vivo* suppression via EF-Tu variants with Sep-OTS (green) or without SepRS (blue) as measured by synthesis of CAT (measured by IC_{50} value). P values relative to EF-coli.

Data generated by the EF-Tu library presented an opportunity to evaluate how effectively REAP identified sites that enhanced promiscuity, especially residue D216. Site 216 was ranked within the top ten sites identified by Type II functional divergence and within the top 15% (out of 279 total residues) overall. Six sites selected for library development were ranked higher: V276, N273, S65, G275, F261, and E259, listed in order ranked. Of these residues, site N273 was also mutated in variant EF-4A. Although EF-N273A was not as effective as EF-D216A or EF-4A, contrary to previous findings N273 does have some impact in promiscuity, evidenced by additional data (Figure 5.3).[52] As expected, while some residues identified by REAP do affect substrate compatibility, amino acid conservation is governed by a variety of factors (e.g., tertiary structure or solubility). Not all residues identified by REAP participate in protein function despite exhibiting strong signatures of functional divergence.

Upon reviewing the data more closely, it is relevant that despite screening more sequence space, these results parallel previous work implicating position 216 in efforts to broaden substrate compatibility.[16] This may be a consequence of using distance discrimination, the prevailing strategy to engineer EF-Tu substrate affinity, to cull the REAP data. Although researchers routinely use external information (e.g., crystal structures) to guide amino acid selection, in this case being influenced by prior knowledge led to a conclusion similar to previous work. This observation suggests that part of REAP's power lies in identifying residues not previously expected to contribute to function. Twenty-two residues, while more than earlier efforts, may have not been a sufficiently diverse group to examine.

5.2.2 Mass spectrometry confirms substrate-promiscuous EF-Tu variants

Table 5.2 A representative group of peptide spectra matched. A modified CAT112TAG gene was translated via EF-Tu variants EF-4A, EF-D216A, and EF-Sep. A control wild-type CAT gene was also translated via EF-Sep and EF-coli.

Protein sequence used to generate CAT protein		CAT112TAG			wild-type CAT	
Annotated Sequence	Modifications	EF-4A	EF-D216A	EF-Sep	EF-Sep	EF-coli
[E].YHDDFRQF.[L]					8	12
[F].SSLWSEYHDDF.[R]					2	
[F].SSLWSEYHDDFRQF.[L]					11	11
[W].SEYHDDFRQF.[L]					17	20
[W].SEYHDDFRQFL.[H]					9	6
[Y].HDDFRQF.[L]					41	19
[Y].HDDFRQFL.[H]					2	
[F].HEQTETFSSLWSEYHDsFRQF.[L]	S17(Phospho)					
[F].SSLWSEYHDsFRQF.[L]	S10(Phospho)	2	2	9		
[F].SSLWSEYHDsFRQFL.[H]	S10(Phospho)					
[W].SEYHDsFRQF.[L]	S6(Phospho)	3		29		
[W].SEYHDsFRQFL.[H]	S6(Phospho)			6		
[F].HEQTETFSSLWSEYHDSFRQF.[L]						
[F].SSLWSEYHDSFRQF.[L]		3	3			
[W].SEYHDSFRQF.[L]						
[W].SEYHDQFRQF.[L]			1			

Electrospray ionization mass spectrometry (chymotrypsin digestion) confirmed site-specific Sep incorporation. Results showed EF-4A and EF-D216A not only incorporated Sep but also serine at the permissive position. EF-D216A also incorporated Gln-tRNA^{Sep} (Table 5.2). Serine and glutamine misincorporation at the amber codon could be explained either by a mischarged tRNA^{Sep} being delivered to the ribosome or by a native aa-tRNA mispairing with the amber codon. If EF-Tu variants EF-4A and EF-D216A were incorporating canonical amino acids via a misacylated aa-tRNA^{Sep}, it would indicate these EF-Tu variants are truly substrate-promiscuous, accepting both ncAA-acylated tRNAs as well as misacylated tRNAs. To confirm, we pursued both possible explanations for the mixed protein product. It is relevant to note, experimental controls reject the possibility of Sep post-translational dephosphorylation as a route to Ser^{112} incorporation. If a post-

translational modification were responsible, Ser^{112} would be evident in the EF-Sep expressed CAT protein as well.

Table 5.3 IC50 values. EF-Tu variants assayed without the SepRS gene. P-values are relative to EF-coli assayed without the SepRS gene.

EF-Tu variant	n (replicates)	mean	st dev	p-value
EF-K263A	3	2.08	2.65	0.000518
EF-Sep	3	13.25	4.5	0.039082

To determine whether a mischarged $tRNA^{Sep}$ or mispaired aa-tRNAs was responsible for serine incorporation, $tRNA^{Sep}$ and SepRS were withheld from translation in separate experiments. When SepRS was withheld, both EF-4A and EF-D216A were able to express a catalytically active target protein (Figure 5.3). Other variants showed low translation without SepRS (Table 5.3). This suggests $tRNA^{Sep}$ has cross-compatibility with native aaRSs and is being misacylated with serine. Supporting this conclusion, when $tRNA^{Sep}$ was withheld from the assay, no CAT expression was observed (5 μg/mL was the lowest concentration tested). This indicates a misacylated $tRNA^{Sep}$ is the mechanism by which serine is being misincorporated.

These results confirmed variants EF-4A and EF-D216A are, in fact, substrate-promiscuous. In contrast to wild type EF-Tu and EF-Sep, both REAP-developed variants will accept misacylated tRNAs and ncaa-tRNAs. This lack of discrimination suggests these EF-Tu variants could be promising in combination with other ncAAs. It also recommends them as platforms for development of additional EF-Tus with novel function, in the form

of either further enhanced promiscuity or alternate substrate specificity.[40] Additionally, these data also emphasize the power of single mutations to EF-Tu. The REAP library was originally developed with each variant containing multiple mutations, a common strategy; however, these data suggest designing EF-Tu variants with single point mutations might be a more valuable strategy especially when seeking broad substrate compatibility.[16]

While these data illustrate the development of promiscuous EF-Tu, it also introduces a possible objection to using a promiscuous EF-Tu: inaccurate translation of the target gene. While this may initially seem problematic, we argue this challenge is readily overcome by recent advances to OTS engineering. Our data confirmed this particular o-tRNA ($tRNA^{Sep}$) is cross-compatible with endogenous aaRSs.[17] This indicates that misincorporation, while permitted by the EF-Tu, is actually caused by a cross-reactive OTS. While a substrate-specific EF-Tu variant, like EF-Sep, can prevent misincorporation, this strategy merely shifts the burden of accurate translation from the aaRS:tRNA pair to the EF-Tu. Additionally, it also fails to address the cross-compatible OTS as the underlying cause.

Unfavorable consequences of cross-compatible OTSs (e.g., depressed host organism growth, inefficient multi-site ncAA incorporation) are well established.[10] Since cross-reactive OTSs are a common obstacle to genetic expansion, recent articles advocate for more rigorous o-tRNAs and o-aaRS engineering.[10, 11, 53, 54] Precise OTS engineering transfers the responsibility of accurate translation from the EF-Tu back to the aaRS:tRNA pair. This redistribution of labor mimics the design of cellular translation. In the cell, endogenous aaRS:tRNA pairs are unequivocally orthogonal to all other tRNAs and aaRSs. While wild type EF-Tu serves as a component of quality control, in point of fact, tRNA

misacylation rarely occurs.[55] Consequently, the primary responsibility for accuracy during translation falls to the canonical aaRS:tRNA pairs, not downstream translation components. As such, it is not surprising that improving OTS orthogonality improves organismal growth and multi-site ncAA translation.[10, 11]

By mirroring the native distribution of responsibilities, discrete OTSs that ensure accurate tRNA acylation pave the way for researchers to use translation components with expanded capabilities, including substrate-promiscuous EF-Tus. Were a rigorously engineered OTS used, there is no evidence a promiscuous EF-Tu would undermine accurate translation. Similarly, we would anticipate that synthetic acylation methods (e.g., flexizymes) to be compatible with a promiscuous EF-Tu.[50]

5.2.3 Selected promiscuous EF-Tus improve organismal fitness

While broad substrate acceptance is desirable in EF-Tu, there may be a limit to the degree of infidelity that it is possible for the cell to tolerate. A promiscuous EF-Tu must be sufficiently indiscriminate to readily accept ncaa-tRNAs; however, too promiscuous and it could interfere with translation of canonical amino acids. Were an engineered EF-Tu sufficiently competitive with native EF-Tu, the translation machinery may inaccurately translate genomic DNA. To examine the impact of a promiscuous EF-Tu on endogenous translation, we compared the effect of EF-Tu variants on organismal fitness.

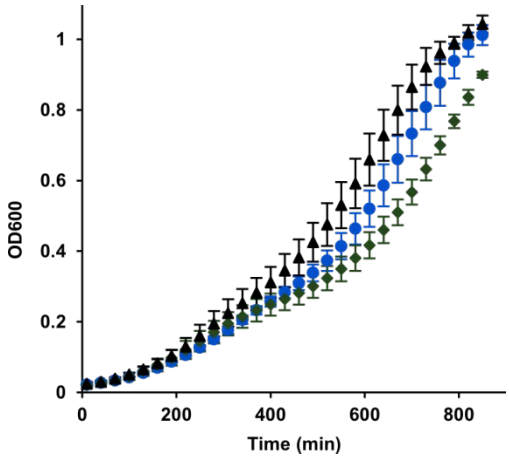

Figure 5.4 Growth assays for EF-Tu variants expressed in BL21ΔserB cell line. Triplicate averages are shown for EF-4A (black triangles), EF-Sep (blue circles), and EF-coli (green diamonds).

Table 5.4 Calculated parameters from growth curves fitted with a modified logistic growth model. Maximum specific growth rate was calculated based on x=t0. Lag time was calculated based on the line tangent to the fitted curve at x=t0.

	Maximum specific growth rate (μ_{max}) [10^{-4} min^{-1}]		Lag time (λ) [min]	
	Best fit value	**Standard deviation**	**Best fit value**	**Standard deviation**
EF-4A	23.8	1.2	329.6	44.0
EF-Sep	25.7	0.6	414.2	19.8
EF-coli	22.9	2.2	450.8	38.9

In order to evaluate the effect of a promiscuous EF-Tu on organismal fitness, we compared growth of cell line BL21Δ*serB* when expressing EF-coli, substrate-specific EF-Sep, and the promiscuous variant EF-4A (Figure 5.4). Cultures expressing the substrate-promiscuous EF-4A grew fastest with a shorter lag time before reaching an exponential growth phase, but both EF-Sep and EF-4A grew noticeably faster than cells expressing EF-coli (Table 5.4). Parameters were calculated based on growth curves fitted with a modified logistic growth model that combined linear models for lower and upper regions of data (assuming an initial lag period and eventual asymptote) with a logistic model y =

1/(1+e^((t0-x)*A)) for the growth phase. In total, seven parameters were fit to each growth curve. Because the BL21Δ*serB* cell line lacks the gene encoding Sep phosphatase, growth curves could not be reproduced using an alternative engineered cell line (e.g., one in which the amber codon has been reassigned).

To some extent, native EF-Tu is both substrate-specific and promiscuous. Because EF-Tu has proof-reading capabilities, by limiting its ability to check for misacylated aa-tRNAs, a promiscuous EF-Tu may increase the rate of misincorporation of canonical amino acids. The cell's translation apparatus has been finely tuned for unparalleled accuracy and speed. As such, expanding the capabilities of this carefully balanced system may have consequences impacting genome translation manifested as depressed organismal growth.

However, unlike tRNAs and aaRSs, EF-Tu already exhibits a certain level of promiscuity. EF-Tu accepts all twenty canonical aa-tRNAs as substrates, which suggests expanding the substrate compatibility of EF-Tu may be within the natural limits of the translation apparatus.

The growth curves indicate that a promiscuous EF-Tu does not negatively impact endogenous translation; in fact, it can actually decrease the fitness cost of expanding the genetic code. These data also indicate a promiscuous EF-Tu does not interfere with translation of the genome. Any misincorporation would have to be minimal so as to not impact host organism fitness. Hence, with evidence of the benefits, including reduced fitness cost, a promiscuous EF-Tu appears to be an asset to genetic code expansion with minimal drawbacks.

5.3 Materials and methods

5.3.1 REAP library

The REAP alignment was generated using thirty-eight sequences from nineteen species of bacteria that express both EF-Tu and SelB. Due to a large discrepancy in average sequence length between EF-Tu and SelB sequences, we normalized the length of the thirty-eight selected sequences to create a more accurate phylogeny. Generally, twenty-five residues were eliminated from the N-terminus of each SelB sequence and 343 residues were deleted from the C-terminus. For EF-Tu sequences, forty-seven were removed from the N-terminus; the C-terminus was not adjusted.

REAP analysis was completed using DIVERGE2.0 software. The multiple sequence alignment was generated in Clustal Omega; the phylogeny was generated within DIVERGE2.0 (using a Poisson distribution). Output was calculated for Gu99, Gu01, and Type II. Residues that were cut from the N-terminus were evaluated by eye.

5.3.2 In vivo assay

Variants were assayed using a system for the co-translational insertion of Sep *in vivo*. This system included an orthogonal triplet, a tRNA ($tRNA^{Sep)}$, an aminoacyl-$tRNA^{Sep}$ synthetase (SepRS), and EF-Tu variant (EFSep) specifically engineered for Sep. These genes were located on two plasmids: pCAT112TAG-SepT (Addgene, plasmid number 34624), and pKD-SepRS-EFSep (Addgene, plasmid number 34623). The EF-Tu variant EFSep was used as a positive control and the standard to which the REAP variants were compared. Wild type *E. coli* EF-Tu, which is not compatible with Sep, was used as a negative control. It is relevant to note that all experiments contained endogenous wild type EF-Tu. Cell lines BL21Δ*serB* (Addgene, bacterial strain number 34929) and Top10Δ*serB*

(Addgene, bacterial strain number 34928) were used. All plasmids and cell lines from Addgene were gifts from Jesse Rinehart and Dieter Söll.

5.3.3 Calculate IC_{50}

Plasmids of choice (pKD and pCAT) were transformed into BL21ΔserB competent cells (Addgene, catalog number 34929). A single colony was selected, grown overnight, and made into glycerol freezer stocks (25% sterile glycerol, 25% sterile water, 50% bacteria culture). For each assay, glycerol freezer stocks were streaked out, a single colony was picked and grown ~24 hours. The culture was then diluted to OD_{600} 0.15 in media supplemented with 2mM Sep, grown to OD_{600} 0.6-0.8 and induced (0.5mM IPTG, Sigma-Aldrich). Cultures were allowed to express for 20 hours then diluted in saline and plated, in duplicate, on petri dishes with a range of chloramphenicol (Thermo Fisher Scientific) concentrations. Colonies were counted daily.

All liquid and solid cultures were grown at 30 ˚C. All liquid cultures were grown in LB media supplemented with 0.08% glucose. Kanamycin (25μg/mL, kanamycin sulfate, VWR), tetracycline (10 μg/mL, tetracycline Hydrochloride 98%, Alfa Aesar), and Sep (2 mM, O-phospho-L-serine Sigma-Aldrich) were present in all liquid cultures and agar Petri dishes.

5.3.4 Protein purification for mass spectrometry

For mass spectrometry assays, a hexahistidine tag was added to the carboxyl-terminus of the CAT112TAG gene (via Gibson assembly). The His-tag was added to the carboxyl-terminus to prevent truncated peptides from being purified. Appropriate glycerol

freezer stocks were made as described above. Glycerol freezer stocks were streaked out, a single colony was picked and grown overnight. Then, 1-1.5 mL starter culture was added to 0.5-3L media supplemented with 2mM Sep, grown to OD 0.6-0.8 and induced (0.5mM IPTG). Protein was expressed for 20 hours then spun down and frozen at -80 ˚C. Cultures were resuspended in 5mL BugBuster® (Protein Extraction Reagent EMD Millipore) and 2.5 uL Benzeonase (250 U/uL purity >90% EMD Millipore) per 1 g cell pellet. Resuspended pullets were incubated, rocking, at room temperature for 60 minutes then spun down (11,419 x g). The supernatant for each sample was collected and applied to a Ni-NTA Superflow Column 12 x 1.5mL (Qiagen) using a vacuum manifold (QIAvac 24 Plus, Qiagen). All filter sterilized buffers contained 50mM NaH2PO4 300mM NaCl pH 8.0 with either 10, 20, or 500 mM imidazole. Columns were prepped by decanting the storage buffer, then applying 10mL 10 mM imidazole buffer. Next, 30mL of supernatant were applied to column, followed by 10mL of 20mM imidazole buffer. This step was repeated, applying 30mL supernatant followed by 10mL 20mM imidazole buffer, until all supernatant had been applied to the column, ending with 10mL 20mM imidazole buffer. Finally, protein was eluted in 0.5mL aliquots of 500mM imidazole buffer (4.5mL total). Eluate aliquots were run on an SDS-PAGE gel to estimate protein concentration in aliquots. When deemed necessary, aliquots were combined and concentrated using Spin-X UF 500uL Centrifugal Concentrator, 10,000MWCO membrane (Corning) .

Purified samples were digested with chymotrypsin and analyzed using ESI-MS (Q Exactive Plus, Thermo Scientific) with a Dionex nano-LC system in front to separate peptides. The wildtype sequence of CAT was also expressed and the amino acid sequence checked via mass spectrometry; as expected, only the wild type amino acid, aspartic acid,

was translated at position 112. Variants EF-coli, EF-N63A, EF-K263A, and EF-N273A were not analyzed using mass spectrometry because protein expression levels were too low to isolate purified CAT protein.

5.3.5 Growth curves

For each sample, glycerol freezer stocks were streaked out. Three colonies were selected from each plate and grown ~24 hours. The starter cultures were then diluted to OD_{600} 0.15 in 200μL fresh media supplemented with 2mM Sep and 0.5mM IPTG. Samples were grown overnight with shaking (medium) in 96-well plates (black sides, clear bottom; Greiner Bio-One, catalog number 655096). A BioTek™ Synergy™ H4 Hybrid microplate reader (Thermo Fisher Scientific) measured absorbance (OD_{600}) with pathlength correction every ten minutes.

Three wells with only LB supplemented with glucose, antibiotics, Sep and IPTG served references for absorbance measurements. All liquid cultures were grown at 30 ˚C in LB media supplemented with 0.08% glucose, 25 μg/mL kanamycin, and 10 μg/mL tetracycline.

5.4 **Additional data**

5.4.1 In vivo assay design

The *in vivo* assay tested our EF-Tu variants compatibility with the ncAA, Sep. *In vivo* translation of the Sep in E. coli was accomplished by using an orthogonal triplet: a tRNA ($tRNA^{Sep}$), synthetase (SepRS), and EF-Tu variant (EFSep) specific for the ncAA.[17] This particular ncAA was not compatible with wildtype E. coli EF-Tu but translation was

accomplished when an EF-Tu variant was engineered to accept Sep as a substrate. We purchased the necessary components, cell lines Top10Δ*serB* and BL21Δ*serB* as well as plasmids pKD-SepRS-EFSep and pCAT112TAG-SepT, to assess our EF-Tu libraries' compatibility with this ncAA. These were a gift from the Soll and Rinehart labs (Addgene). EFSep was used as a positive control and the standard to which our variants were compared. Wild type *E. coli* EF-Tu, which is not compatible with Sep, was used as a negative control.

Both SepRS and EFSep genes on the pKD-SepRS-EFSep plasmid (Kan resistant, grows at 30 deg.) are controlled by the lac promoter and expression can be induced via IPTG. The pCAT112TAG-SepT plasmid (Tet resistant) has the CAT gene with an amber stop codon at the permissive position, Asp^{112}, along with the $tRNA^{Sep}$ (SepT). When Sep is translated, the CAT gene is expressed and imparts chloramphenicol resistance to the bacteria, meaning an IC_{50} can be calculated and directly related to the efficiency with which Sep is incorporated. All assays were performed in BL21Δ*serB* cells.

5.4.2 Cloning EF-Tu and tRNA variants

We excised the EFSep gene and cloned EF-Tu variants from our ASR, consensus, REAP, and transplantation libraries into the pKD-SepRS plasmid using a restriction enzyme digestion and ligation and confirmed via sequencing. Dr. Cole ordered primers to add relevant digestion sites to the EF-Tu genes via PCR and we successfully cloned all EF-Tu variants from the ASR, REAP, and consensus libraries into the pKD-SepRS plasmid for use with the *in vivo* assay. The first ligation was not successful. We ran a serious of controls and confirmed that the ligation reagents were active. We found a one to one molar ratio of insert to vector resulted in successful ligations. It is worthy of note that one ASR

variant, 209, was not carried through the *in vivo* assay because it contains one of the digestion sites in the middle of the gene.

We also mutated $tRNA^{Sep}$ gene based on unpublished work (Gaucher Lab) that indicates the position of a G-U wobble pair affects the strength of aa-tRNA:EF-Tu association. Although optimization of the PCR annealing temperature and transformation protocol was time-consuming, Agilent's QuikChange II Site-Directed Mutagenesis kit was used to generate these variants. It is worth noting that both NovaBlue and XL1-Blue supercompetent cells are Tet resistant and thus incompatible with the pCAT plasmid. We moved a G-U wobble pair to various positions in the T-arm and acceptor stem of the tRNA to perturb the strength of Sep-$tRNA^{Sep}$:EF-Tu association. We also made one variant that lacked the $tRNA^{Sep}$ as a control.

During this process we learned that although these plasmids are high copy, minipreps resulted in low plasmid concentration which often resulted in poor sequencing reads. As such, we used larger culture volumes for minipreps and also started submitting higher plasmid DNA concentrations for sequencing. Primer concentrations were also optimized for sequencing submissions.

5.4.3 Assay implementation

In preparation for assays, we made chemically competent cells of both Top10Δ*serB* and BL21Δ*serB* cell lines. BL21Δ*serB* chemically competent cells containing plasmids (either pKD-SepRS-EFSep or pCAT112TAG-SepT) were also made. Stocks of Tet, Kan, Cmp, and Sep. All stocks were filter sterilized (0.22um) except the Sep stock as the ncAA has low solubility in water and would have been filtered out of solution. Contamination from Sep was never an issue.

The published methods were incomplete. By contacting the Rinehart Lab, over three months we learned that 1) they use the BL21Δ*serB* cell line for expression, not Top10Δ*serB* as stated in the publication, 2) they expressed cultures for about 20 hours at 30 degrees, and 3) they used LB liquid media supplemented with 0.08% glucose. Along the way, we learned to omit IPTG in the assay plates. Through trial and error we learned the IPTG concentration required for the plate assay (unreported) was roughly ten times than what was reported for phospho-protein expression (0.05 and 0.1 mM).

Another set of challenges with developing a working protocol related to how slowly these cultures grew. First, these cultures did not reach confluency in the 12-18 hours a typical overnight cultures would take. Our cultures took about 24-30 hours to grow to confluency. Second, overnight starter cultures worked best if diluted to only 0.15 OD_{600} whereas we usually diluted to roughly 0.001 OD_{600} in other expression protocols. Third, control plates took 30-48 hours to grow and assay plates took four to seven days even with low Cmp concentrations. As a reference point, doubling time in liquid media was roughly 70 minutes but depending on the variant could be over 100 minutes. These extended growth times meant that we did not think the protocol was working when in fact we just weren't waiting long enough for the cultures to grow.

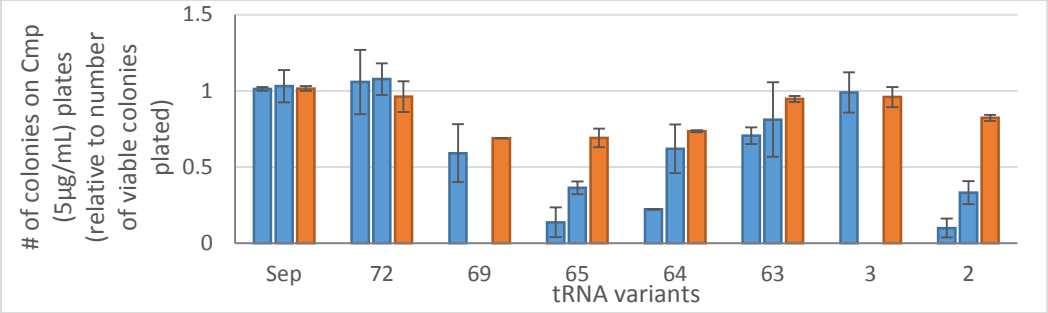

Figure 5.5 Growth of tRNASep variants on Cmp plates (5 µg/mL). Data is average number of colonies on two plates (same liquid culture). Cultures diluted in LB (blue) or saline (orange) before plating. Counted Day 3.

Reproducibility was also a significant problem. Eventually, the Rinehart and Söll Labs released an updated phospho-protein kit. While this kit was not relevant to assay EF-Tus, it had an in-depth protocol which provided insight into the methods these labs used to develop their assays. We not only learned how slowly these bacteria strains grow, but also that reproducibility was a significant problem. The revised protocol recommended making glycerol freezer stocks of transformations, then streaking out from these stocks for each assay. In typical practice, a fresh transformation results in improved protein expression; using glycerol stocks significantly improved reproducibility.

Our first goal was to translate Sep and replicate the published data. We optimized growth temperatures, induction conditions, IPTG concentration, expression time, and plating and dilution procedures. Following recommendations from the developing lab, we made glycerol freezer stocks from overnight cultures grown from fresh transformations. Glycerol stocks contained a 1:1 ratio of overnight bacteria culture and sterile 50% glycerol solution. These glycerol freezer stocks were streaked out and grown 30 hours then colonies were picked to assay.

Chronologically, the first challenge with the *in vivo* assay was that the published methods section which accompanied the phospho-protein translation kit (Addgene) was not sufficient to reproduce the published data nor to synthesize phospho-proteins. The Torres Lab (Georgia Tech) also struggled with this kit. As such, the protocol used was developed in-house; we optimized growth temperature, cell line selection, induction procedure, expression times, reproducibility, and plating and dilution technique. It is relevant to note the bacteria strains used grow slowly (in both liquid and plated cultures). As such, we also assayed addition protein libraries including ASR (12 variants), consensus (2 variants), and the binding-site transplantation library (15 variants) although these data did not result in a publication.

The following data were generated from efforts to replicate published results and improve reproducibility. Since data is reported at IC_{50} values, we needed a reliable technique to dilute and plate bacteria cultures following induction. We tried diluting in LB and saline and found that diluting in saline offered significantly better reproducibility (Figure 5.5).

We also tried to increase antibiotic concentrations (Figure 5.6). Several people in lab were observing tiny colonies on plates. It seems possible now that we had bacteriophage contamination but at the time increasing Kan was one attempted strategy. What we learned as that doubling the concentration of Kan seemed to improve growth, but did not affect the tiny colonies either. Eventually, we were able to eliminate the contamination issue, but increasing antibiotic concentrations did not help.

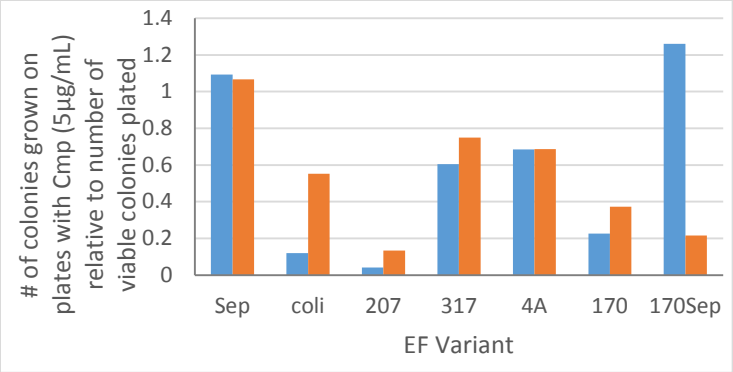

Figure 5.6 Testing increasing antibiotic concentration. Increased kanamycin concentration from 25 µg/mL (blue bars) to 50 µg/mL (orange bars). Data collected on day 3.

We also examined the concentration of IPTG used to induce gene expression (Figure 5.7). We found it did not make much difference so we continued to use 0.5mM IPTG.

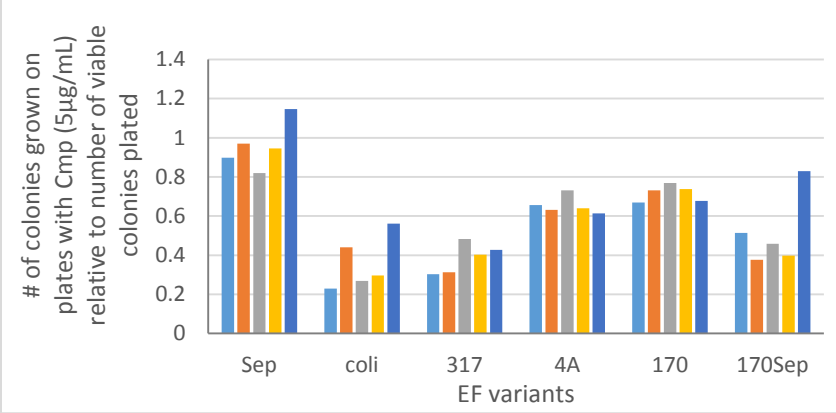

Figure 5.7 Testing concentrations of IPTG. Concentrations tested included 0.1mM (light blue), 0.5mM (orange), 0.75mM (gray), 1mM (yellow), 5mM (navy). Data collected after 2 days of growth.

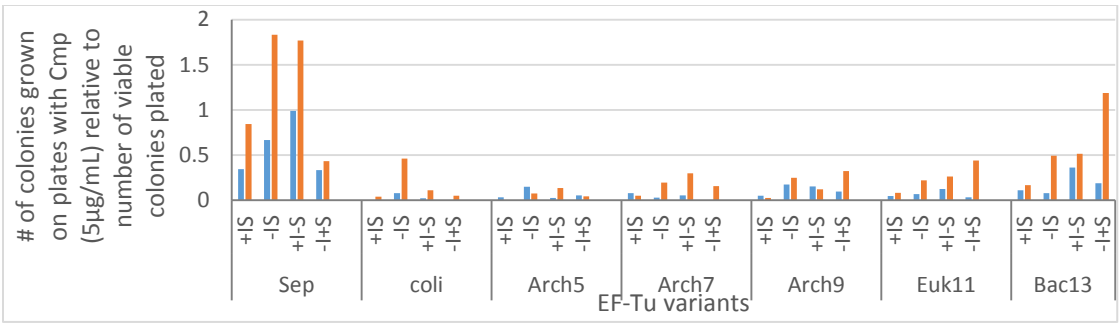

Figure 5.8 Trying to measure background. Samples contained (+) or lacked (-) IPTG (I) and phosphoserine (S). Samples were plated on plates containing phosphoserine (orange bars) and lacking phosphoserine (blue bars). Data collected on day 2.

We also attempted to run a series of controls with or without the ncAA (Sep) and with or without IPTG (Figure 5.8). Growth was consistently higher when Sep was added to plates (orange bars). As such, we decided to always add Sep to assay plates. These data also indicate our operon is leaky which is not especially surprising. What was surprising was that growth rates increased when Sep was withheld from the expression culture suggesting that endogenous Sep is more bioavailable and that Sep up-take was somehow detrimental to growth. However, because ncAAs are most commonly added to liquid culture prior to induction, we continued with that method. We also continued to induce with IPTG.

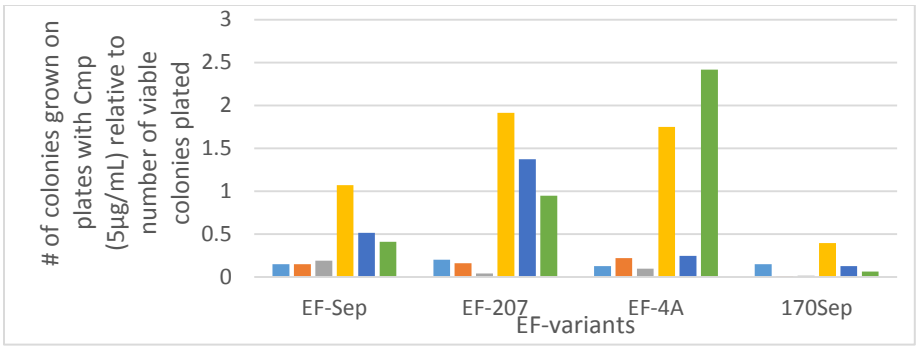

Figure 5.9 Varying Sep concentration. Samples with Sep only in liquid media (not plates) contained 0.5 (blue), 2 (orange), or 4 (gray) mM Sep. Reactions with Sep in both liquid and solid media contained 0.5 (yellow), 2 (navy), 4 (green) mM Sep.

We also tried different Sep concentrations in both liquid LB media and solid agar plates. These results confirm that having Sep in the plates greatly improved growth rates, but they do not consistently suggest a specific Sep concentration is most desirable. Most data suggests 0.5 mM is the ideal concentration but EF-4A (which we are trying to optimize) suggests 4 mM is better. Other articles suggest 1 mM is best. We decided to continue to use 2 mM which is recommended by the article (Figure 5.9).

Still trying to replicate published data, we tested EF-Sep and EF-coli (Figure 5.10). EF-coli, our negative control, is in keeping with published values but EF-Sep, a standard for comparison, is lower than reported.[17] Eventually, we were able to increase EF-Sep IC_{50} by greatly increasing the concentration of IPTG used but it minimized the difference between our REAP-variants and negative control.

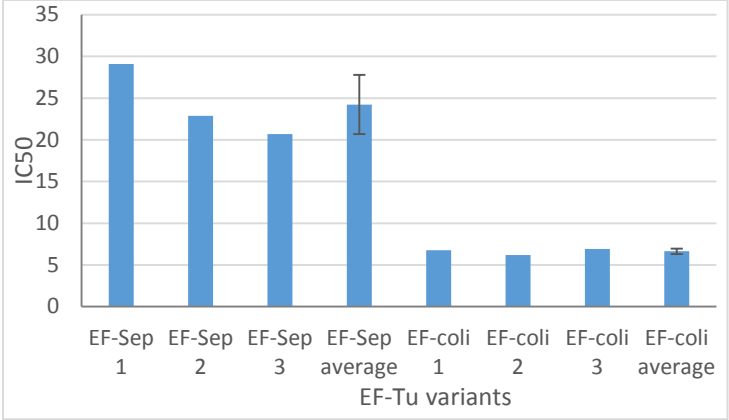

Figure 5.10 IC_{50} of positive control, phosphoserine, (blue bars) and negative control, coli, (orange bars). Phosphoserine data from Day 6. Coli data from Day 7.

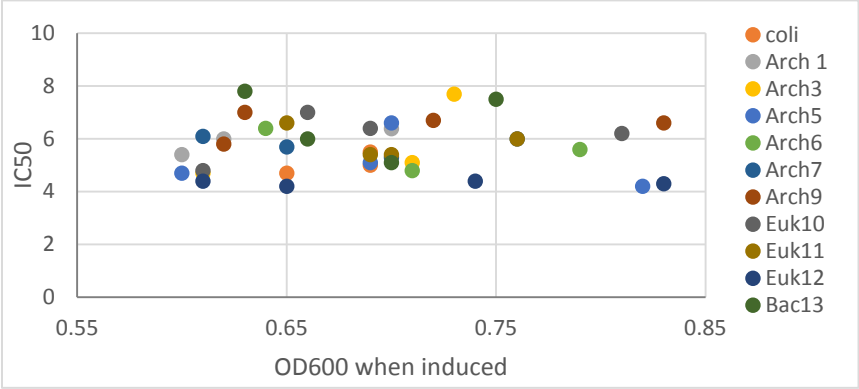

Figure 5.11 OD_{600} when induced versus IC_{50} for binding-site library. Assays to determine if there is an ideal OD_{600} at which induction would produce a higher IC_{50}.

We also tried to optimize our OD_{600} at induction, but our system did not seem to be significantly sensitive (Figure 5.11). There was no obvious correlation between OD_{600} and the IC_{50}. At this point, we decided to progress with assaying the EF-Tu variants. The binding-site transplantation library showed no significant hits (Figure 5.12). We did not find any hits in the ASR library (Figure 5.13). However, we had a hit, EF-4A, in the REAP library (Figure 5.14). We pursued this variants to write the article. Following identification

of EF-4A as a possible ncAA-compatible variant. We returned to the *in vitro* assay (see Chapter 4) but were unable to generate results.

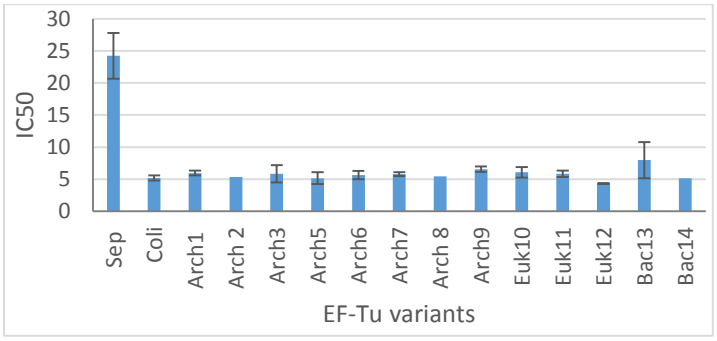

Figure 5.12 IC_{50} of EF-Tu variants from binding-site library. Error bars are standard deviation. Data collected day5.

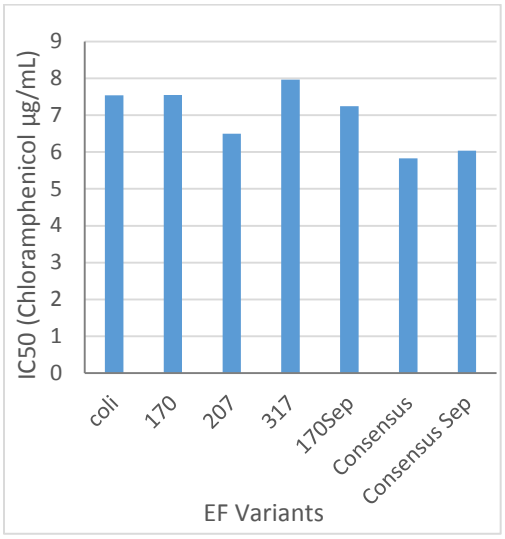

Figure 5.13 IC_{50} of select ancestral and consensus EF variants. These variants showed the most promise in previous studies. EF coli is the negative control. Data collected on day 10.

Following our *in vitro* efforts, we were requested to try to expand the *in vivo* assay to assay EF-4A with other ncAAs. We agreed to work for three months trying to expand the *in vivo* work in such a way as to assay EF4A *in vivo* against tyrosine analogs. The new target of the EF-Tu project was to be aromatic rings and tyrosine analogs. We used

aaRS:tRNA pairs developed by the Schultz lab. We began using SOE to make the genes and Gibson assembly to clone them. Unfortunately, although eventually we were able to make these techniques work, starting them up in lab from scratch took time. The plasmids we worked with replicated slowly, grew at lower temperatures, and the first two or three protocols developed for Gibson assembly failed. By the time trouble shooting was done, more than three months had passed and we were eager to move back to the original EF4A data and try to write the relevant article. However, first we were requested to initiate the Gaucher Lab's involvement in the Multidisciplinary University Research Initiative (MURI) and train a new postdoctoral fellow for that grant (see Chapter 6).

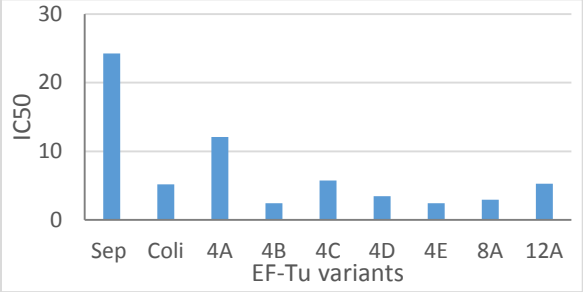

Figure 5.14 IC_{50} of REAP EF-Tu variants. Data is preliminary. Variant 4F not shown. Data collected day 6.

CHAPTER 6. INITIATING MURI GRANT IN LAB

We initiated research on a Multidisciplinary University Research Initiative (MURI), from the Department of Defense including transfer of materials and communication of results via video conference calls. The aim of this project is to develop a fully engineered translation system which can accommodate a wide range of monomers for applications in synthetic polymer synthesis.

We successfully introduced a new assay in the lab which allowed analysis of our EF-Tu variants with OTSs from collaborating labs. We initiated transfer of materials from collaborating labs. During and after transfer of materials, we initiated email chains and conference calls and introduced new lab members to collaborators as appropriate.

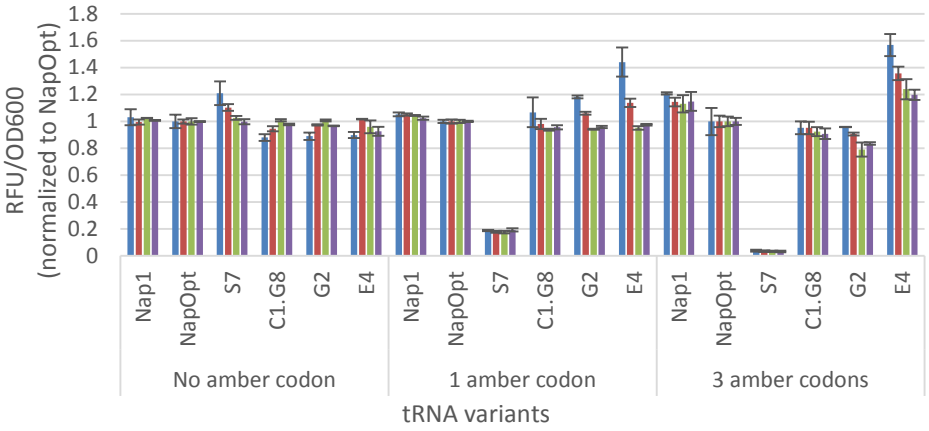

Figure 6.1 Replicating published data expressing GFP with tRNA variants (in duplicate). Solutions diluted 1:8 (blue), 1:4 (red), 1:2 (green), or not diluted (purple).

Using Gibson assembly, we synthesized the plasmids required to assay our EF-Tus with the Ellington's OTSs. We also ordered materials and arranged to use equipment to assay components. Following synthesis of plasmids, we successfully replicated the Ellington lab's published work (Figure 6.1). We assumed our samples would be too diluted

for the microplate reader to accurately read the OD_{600}, but we found not diluting the samples aligned best with the published data. Measurements were taken ten times and averaged for each well. Duplicate wells were assayed for each sample.

We then moved on to assay our EF-Tus in conjunction with the Ellington lab's engineered aaRSs and tRNAs. Initial results looked promising (Figure 6.2). However, low reproducibility obscured meaning.

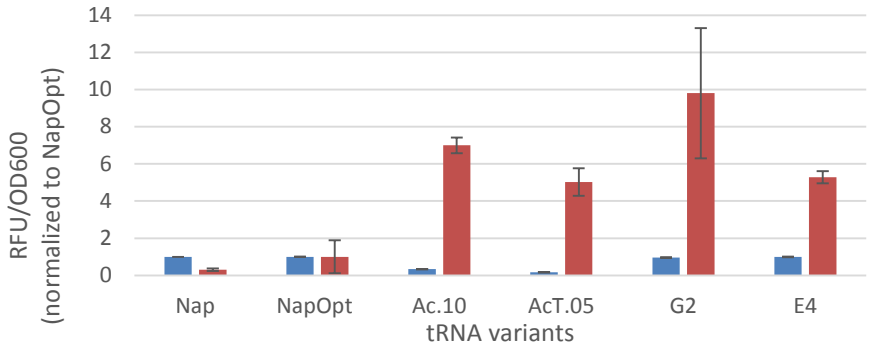

Figure 6.2 Translation of GFP (1 amber mutation) by six tRNA variants. Results in triplicate show translation with no added EF-Tu (blue) and EF-4A (red).

We began trouble-shooting problems with data reproducibility and decided the origin of replication on the EF-Tu plasmids was not compatible with the plasmids from the Ellington lab. At this point, we trained and transferred this aspect of the project to a postdoctoral fellow.

CHAPTER 7. FUTURE DIRECTIONS

One challenge associated with this project was the size of the libraries. All libraries were extremely small (i.e. under 20 variants). With libraries that small, it becomes debatable if it is, in fact, a library or if it is a collection of very specifically engineered protein variants. Setting diction aside, these libraries were so small even with a highly functional library there was still a somewhat low probability of observing any desired outcome. Even previously published data using the REAP method used a library nearly an order of magnitude larger (over 90 variants).[56] The difference in library size tentatively suggests that a larger library or a different tactic might be necessary to identify a variant with the desired activity.

A second challenge was developing a functional assay. Despite all our efforts, the *in vitro* assay consistently suffered from reliability and reproducibility. Regularly the control experiments would stop working and required weeks to months of expensive and time-consuming work to identify the problem. *In vitro* translation is well known to be a challenging technique to master. Additionally, the radioactive label was highly temperature sensitive and degraded rapidly. It also had alanine and valine contamination. Eventually, we found a thermostable radio-compound, but we learned to be more vigilant about independently evaluating an inherited protocol.

The *in vivo* assay was also impractical due to the timescale. The assay took ten to fifteen days to run one assay. In addition, the amount of time required to monitor samples

greatly limited the number of experiments that could be run concurrently or even in parallel.

Even if these assays had been functional, an extremely low-throughput assay is, in hind-sight, potentially problematic given the EF-Tu libraries. With such small libraries, the probability of finding a hit was lower (even with a highly functional library). An alternative would have been to work with a high-throughput library. In fact, once we started working in 96-well plates that could be read on a plate reader, we began to collect relevant data. A high-throughput assay would have also allowed us to expand the variables we worked with. For example, instead of expanding the protein library, we could have utilized a wider range of other translation components (aaRS, tRNAs, ribosomes, etc.) which, when working with an extremely limited number of variants, is another tactic to expand the data gathered and improve the probably of getting a hit.

Another challenge with this project was that utility had to be proven; a higher bar than discovery-based research. This was a challenging goal to achieve since in order for results to be worthy of publication, we had to prove our library variants not only worked but were actually better than existing options. This is a much higher standard and much more difficult to achieve.

In recent months, we have been learning about high-throughput strategies from other labs. Given the data generated, we anticipate a larger library would be critical for moving forward. Alanine scanning or site-saturation mutagenesis would both be viable methods. Additionally a suitably complementary high-throughput assay, such as fluorescence activated cell sorting (FACS), would be important to move forward.

The data herein, suggests that REAP can be successfully used to identify sites suitable for mutation. They also suggest a single mutation can expand EF-Tu substrate compatibility. To the best of our knowledge, a comprehensive site-directed analysis of EF-Tu has not been performed. Given the results generated, we theorize that EF-Tu variants with single mutations might be more powerful than testing multiple mutations.

We would recommend developing a larger EF-Tu library by both expanding the number of sites targeted and by expanding the number of amino acids incorporated as substitutions. We recommend using a site-saturation approach via an NNK codon to target sites identified by REAP and ultimately generate a large library of single mutants. We also recommend targeting sites farther from the target. Although proximity to the amino acid binding pocket is a common method to select residues for mutation in EF-Tu, we theorize that an important advantage to using REAP is the ability to identify amino acid residues that might not be typically selected for mutation. We would be interested to see if a certain amino acid can be identified as the best substitution for EF-Tu.

Complementary to this aim, we would also recommend adopting FACS as a high-throughput screening technique. Using sfGFP as a reporter gene with either 1 or 3 amber mutations would permit much larger throughput than *in vitro* or *in vivo* assays.

Tentatively, we would recommend using a two plasmid system. One plasmid would contain the EF-Tu variant and sfGFP. Another plasmid would contain aaRS:tRNA pairs. We would expect to originally screen using a sfGFP with three amber mutations since published data suggests an engineered EF-Tu might be more influential for multi-site ncAA incorporation. We would recommend building the cassettes from scratch rather than using purchased systems because we think it would permit greater control over promoter

selection. A challenge to this plan is that we might need a second fluorophore (mCherry, perhaps) that would allow us to normalize fluorescence to the number of plasmids expressing sfGFP in the cell. Both mCherry and sfGFP are fast maturing fluorophores which would benefit the system. There are obvious details to consider, for example sfGFP has a rather broad emission peak which overlaps with mCherry's excitation spectrum; however, this pair is not uncommon to use jointly and we expect it would be suitable.

Alternatively, an alanine-scanning approach could be used to identify relevant sites for mutation. While it has drawbacks, alanine scanning has been used to elucidate the role of amino acid residues in many proteins. To the best of our knowledge, this has not been attempted with EF-Tu which would put this in the category, perhaps, of fundamental protein research, thus having a lower bar to publication.

The great challenge of working with EF-Tu is that results are predicated on the ncAA used. It is challenging to make sweeping statements about a helpmate protein. This is another reason why a high-throughput assay in which EF-Tu variants could be compared again many ncAAs would be tremendously useful.

An alternative option would be to use post-translational modifications to detect amino acid incorporation. One of the challenges of assaying EF-Tu is the need to assay for broad incorporation. One way to do this would be to specifically incorporate ncAAs that can be acted upon post-translationally. Then, instead of detecting incorporation of the ncAA, you could detect the modified protein. You may even be able to detect incorporation of different substrates if they were acted upon differently post-translationally. A related strategy would be to use orthogonal tRNAs and orthogonal aaRSs that are known to be

polyspecific. By using polyspecific tRNAs and aaRS, we may generate more data and see a larger number of substrates be incorporated at that specific location.

Alternatively, as a completely different approach it could be interesting to modify SelB for expanded substrate capacity. The advantage of this tactic is that SelB recognizes tRNAs that contain an additional nucleotide base pair in the acceptor stem. A modified SelB might be to exclusively recognize a specific subset of tRNAs that contain an extra nucleotide in the acceptor stem. This strategy might allow close control over which orthogonal tRNAs associate with the modified EF-Tu.

REFERENCES

[1] Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code, In *Annu. Rev. Biochem.* (Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., and Thorner, J. W., Eds.), pp 413-444.

[2] Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - Expanding the chemistry in biology, *Chemical Science 6*, 50-69.

[3] Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs, *Chemistry & Biology 8*, 883-890.

[4] Wang, L., and Schultz, P. G. (2005) Expanding the genetic code, *Angewandte Chemie 44*, 34-66.

[5] Des Soye, B. J., Patel, J. R., Isaacs, F. J., and Jewett, M. C. (2015) Repurposing the translation apparatus for synthetic biology, *Curr. Opin. Chem. Biol. 28*, 83-90.

[6] Fan, C. G., Ip, K., and Soll, D. (2016) Expanding the genetic code of Escherichia coli with phosphotyrosine, *FEBS Lett. 590*, 3040-3047.

[7] Ieong, K. W., Pavlov, M. Y., Kwiatkowski, M., Forster, A. C., and Ehrenberg, M. (2012) Inefficient delivery but fast peptide bond formation of unnatural L-aminoacyl-tRNAs in translation, *J. Am. Chem. Soc. 134*, 17955-17962.

[8] Hong, S. H., Ntai, I., Haimovich, A. D., Kelleher, N. L., Isaacs, F. J., and Jewett, M. C. (2014) Cell-free protein synthesis from a release factor 1 efficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation, *ACS Synth. Biol. 3*, 398-409.

[9] Liu, Y., Kim, D. S., and Jewett, M. C. (2017) Repurposing ribosomes for synthetic biology, *Curr. Opin. Chem. Biol. 40*, 87-94.

[10] Maranhao, A. C., and Ellington, A. D. (2017) Evolving orthogonal suppressor tRNAs to incorporate modified amino acids, *ACS Synth. Biol. 6*, 108-119.

[11] Gan, R., Perez, J. G., Carlson, E. D., Ntai, I., Isaacs, F. J., Kelleher, N. L., and Jewett, M. C. (2017) Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins, *Biotechnol. Bioeng. 114*, 1074-1086.

[12] Liu, C. C., Jewett, M. C., Chin, J. W., and Voigt, C. A. (2018) Toward an orthogonal central dogma, *Nat. Chem. Biol. 14*, 103-106.

[13] Amiram, M., Haimovich, A. D., Fan, C. G., Wang, Y. S., Aerni, H. R., Ntai, I., Moonan, D. W., Ma, N. J., Rovner, A. J., Hong, S. H., Kelleher, N. L., Goodman,

A. L., Jewett, M. C., Soll, D., Rinehart, J., and Isaacs, F. J. (2015) Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids, *Nat. Biotechnol. 33*, 1272-+.

[14] Mendel, D., Cornish, V. W., and Schultz, P. G. (1995) Site-directed mutagenesis with an expanded genetic code, *Annu. Rev. Biophys. Biomol. Struct. 24*, 435-462.

[15] Guo, J. T., Melancon, C. E., Lee, H. S., Groff, D., and Schultz, P. G. (2009) Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids, *Angewandte Chemie 48*, 9148-9151.

[16] Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. (2007) Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system, *J. Am. Chem. Soc. 129*, 14458-14462.

[17] Park, H. S., Hohn, M. J., Umehara, T., Guo, L. T., Osborne, E. M., Benner, J., Noren, C. J., Rinehart, J., and Soll, D. (2011) Expanding the genetic code of Escherichia coli with phosphoserine, *Science 333*, 1151-1154.

[18] Ieong, K. W., Pavlov, M. Y., Kwiatkowski, M., Ehrenberg, M., and Forster, A. C. (2014) A tRNA body with high affinity for EF-Tu hastens ribosomal incorporation of unnatural amino acids, *RNA 20*, 632-643.

[19] Duro-Castano, A., Conejos-Sanchez, I., and Vicent, M. J. (2014) Peptide-Based Polymer Therapeutics, *Polymers 6*, 515-551.

[20] Gentilucci, L., De Marco, R., and Cerisoli, L. (2010) Chemical modifications designed to improve peptide stability: Incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization, *Curr. Pharm. Des. 16*, 3185-3203.

[21] Lang, K., and Chin, J. W. (2014) Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins, *Chemical Reviews 114*, 4764-4806.

[22] Stephanopoulos, N., and Francis, M. B. (2011) Choosing an effective protein bioconjugation strategy, *Nature Chemical Biology 7*, 876-884.

[23] Elliott, T. S., Bianco, A., and Chin, J. W. (2014) Genetic code expansion and bioorthogonal labelling enables cell specific proteomics in an animal, *Current Opinion in Chemical Biology 21*, 154-160.

[24] Stephanopoulos, N., Tong, G. J., Hsiao, S. C., and Francis, M. B. (2010) Dual-Surface Modified Virus Capsids for Targeted Delivery of Photodynamic Agents to Cancer Cells, *Acs Nano 4*, 6014-6020.

[25] Yanagisawa, T., Umehara, T., Sakamoto, K., and Yokoyama, S. (2014) Expanded Genetic Code Technologies for Incorporating Modified Lysine at Multiple Sites, *Chembiochem 15*, 2181-2187.

[26] Wallat, J. D., Rose, K. A., and Pokorski, J. K. (2014) Proteins as substrates for controlled radical polymerization, *Polymer Chemistry 5*, 1545-1558.

[27] Merrifie.Rb. (1969) SOLID-PHASE PEPTIDE SYNTHESIS, *Advances in Enzymology and Related Areas of Molecular Biology 32*, 221-&.

[28] Goto, Y., Katoh, T., and Suga, H. (2011) Flexizymes for genetic code reprogramming, *Nat. Protoc. 6*, 779-790.

[29] Smith, M. T., Wilding, K. M., Hunt, J. M., Bennett, A. M., and Bundy, B. C. (2014) The emerging age of cell-free synthetic biology, *Febs Letters 588*, 2755-2761.

[30] Walsh, C. T. (2014) Blurring the Lines between Ribosomal and Nonribosomal Peptide Scaffolds, *Acs Chemical Biology 9*, 1653-1661.

[31] Josephson, K., Hartman, M. C. T., and Szostak, J. W. (2005) Ribosomal synthesis of unnatural peptides, *J. Am. Chem. Soc. 127*, 11727-11735.

[32] Hartman, M. C. T., Josephson, K., Lin, C. W., and Szostak, J. W. (2007) An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides, *PLoS One 2*.

[33] Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity, *Nature Biotechnology 20*, 1044-1048.

[34] LaRiviere, F. J., Wolfson, A. D., and Uhlenbeck, O. C. (2001) Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation, *Science 294*, 165-168.

[35] Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991) SELENOCYSTEINE - THE 21ST AMINO-ACID, *Molecular Microbiology 5*, 515-520.

[36] Cole, M. F., and Gaucher, E. A. (2011) Utilizing natural diversity to evolve protein function: applications towards thermostability, *Current Opinion in Chemical Biology 15*, 399-406.

[37] Gaucher, E. A., Thomson, J. M., Burgan, M. F., and Benner, S. A. (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins, *Nature 425*, 285-288.

[38] Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A., and Sanchez-Ruiz, J. M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of precambrian beta-lactamases, *J. Am. Chem. Soc. 135*, 2899-2902.

[39] Goldsmith, M., and Tawfik, D. S. (2013) Enzyme Engineering by Targeted Libraries, In *Methods in Protein Design* (Keating, A. E., Ed.), pp 257-283.

[40] Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective, In *Annu. Rev. Biochem.* (Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., and Thorner, J. W., Eds.), pp 471-505.

[41] Gaucher, E. A. (2007) Ancestral sequence reconstruction as a tool to understand natural history and guide synthetic biology: realizing and extending the vision of Zuckerkandl and Pauling, In *Ancestral Sequence Reconstruction* (Liberles, D. A., Ed.), pp 20-33, Oxford University Press, New York

[42] Cox, V. E., and Gaucher, E. A. (2014) Engineering proteins by reconstructing evolutionary adaptive paths, In *Directed Evolution Library Creation: Methods and Protocols, 2nd Edition* (Gillam, E. M. J., Copp, J. N., and Ackerley, D. F., Eds.), pp 353-363.

[43] Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., Ortlund, E. A., and Benner, S. A. (2010) Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection, *Proc. Natl. Acad. Sci. U.S.A. 107*, 1948-1953.

[44] Gu, X. (1999) Statistical methods for testing functional divergence after gene duplication, *Mol. Biol. Evol. 16*, 1664-1674.

[45] Gu, X. (2001) Maximum-likelihood approach for gene family evolution under functional divergence, *Mol. Biol. Evol. 18*, 453-464.

[46] Gu, X. (2006) A simple statistical method for estimating Type-II (Cluster-Specific) functional divergence of protein sequences, *Mol. Biol. Evol. 23*, 1937-1945.

[47] Gu, X., and Vander Velden, K. (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family, *Bioinformatics 18*, 500-501.

[48] Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J., and Sachs, J. N. (2012) The methionine-aromatic motif plays a unique role in stabilizing protein structure, *J. Biol. Chem. 287*, 34979-34991.

[49] Stank, A., Kokh, D. B., Fuller, J. C., and Wade, R. C. (2016) Protein Binding Pocket Dynamics, *Acc. Chem. Res. 49*, 809-815.

[50] Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006) A highly flexible tRNA acylation method for non-natural potypeptide synthesis, *Nat. Methods 3*, 357-359.

[51] Hartman, M. C. T., Josephson, K., and Szostak, J. W. (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids, *Proc. Natl. Acad. Sci. U.S.A. 103*, 4356-4361.

[52] Lee, S., Oh, S., Yang, A., Kim, J., Soll, D., Lee, D., and Park, H. S. (2013) A facile strategy for selective incorporation of phosphoserine into histones, *Angewandte Chemie 52*, 5771-5775.

[53] Fan, C. G., Ho, J. M. L., Chirathivat, N., Soll, D., and Wang, Y. S. (2014) Exploring the substrate range of wild-type aminoacyl-tRNA synthetases, *ChemBioChem 15*, 1805-1809.

[54] Kunjapur, A. M., Stork, D. A., Kuru, E., Vargas-Rodriguez, O., Landon, M., Soll, D., and Church, G. M. (2018) Engineering posttranslational proofreading to discriminate nonstandard amino acids, *Proc. Natl. Acad. Sci. U.S.A. 115*, 619-624.

[55] Ibba, M., and Soll, D. (2004) Aminoacyl-tRNAs: setting the limits of the genetic code, *Genes Dev. 18*, 731-738.

[56] Risso, V. A., Manssour-Triedo, F., Delgado-Delgado, A., Arco, R., Barroso-DelJesus, A., Ingles-Prieto, A., Godoy-Ruiz, R., Gavira, J. A., Gaucher, E. A., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2015) Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history, *Mol. Biol. Evol. 32*, 440-455.